OM of: US-09-303-518D-571 to: SPTREMBL_19:* out_format : pfs

Date: Jun 30, 2002 8:14 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142714_4338/app_query.fasta_1.23501
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_1967 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-571
Query length: 894
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 883.700000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_bacteriap:Q9JY07	+	1469.50	2213.15	2.1e-115	298	! Q9jy07 neisseria meningitidis (
sp_bacteriap:Q9JVV7	+	1467.50	2210.13	3.1e-115	298	! Q9jvv7 neisseria meningitidis (
sp_bacteria:Q9EYM0	+	288.00	425.32	7.2e-16	337	! Q9eym0 acinetobacter sp. bd413.
sp_bacteriap:Q9I7B5	+	266.50	394.13	4.5e-14	295	! Q9i7b5 pseudomonas aeruginosa.
sp_bacteriap:Q9PDN1	+	258.00	380.60	2.4e-13	316	! Q9pdn1 xylella fastidiosa. lipi
sp_bacteria:O33437	+	226.00	334.41	1.1e-10	253	! O33437 pseudomonas fluorescens.
sp_bacteriap:Q9HYZ8	+	222.50	327.04	2.3e-10	312	! Q9hyz8 pseudomonas aeruginosa.
sp_bacteriap:Q9JYV2	+	204.00	299.83	8.2e-09	289	! Q9jyv2 neisseria meningitidis (
sp_bacteriap:Q9JTU4	+	202.00	296.80	1.2e-08	289	! Q9jtu4 neisseria meningitidis (
sp_bacteriap:Q9PH42	+	192.00	282.03	8.3e-08	279	! Q9ph42 xylella fastidiosa. lipi
sp_bacteria:Q9API8	+	178.00	260.41	1.3e-06	292	! Q9api8 methylobacterium extorqu
sp_bacteria:Q9L7Q3	+	176.00	256.57	1.9e-06	317	! Q9l7q3 haemophilus ducreyi. htr
sp_bacteria:Q9L7U9	+	176.00	256.54	1.9e-06	318	! Q9l7u9 haemophilus ducreyi. lip
sp_bacteriap:O84013	+	175.00	251.48	2.6e-06	455	! O84013 chlamydia trachomatis. a
sp_bacteriap:Q9PL30	+	162.00	231.86	3.2e-05	453	! Q9pl30 chlamydia muridarum. lip
sp_bacteriap:Q9ZCL1	+	159.00	231.74	5.0e-05	290	! Q9zcl1 rickettsia prowazekii. l
sp_bacteria:O82937	+	156.50	226.62	8.5e-05	332	! O82937 escherichia coli o157:h7
sp_bacteria:Q9ZGU3	+	156.50	226.30	8.6e-05	343	! Q9zgu3 escherichia coli o157:h7
sp_bacteria:Q9AKA6	+	155.50	226.45	9.9e-05	290	! Q9aka6 rickettsia typhi. htrb p
sp_bacteriap:Q9CME5	+	154.00	223.43	0.0001	313	! Q9cme5 pasteurella multocida. h
sp_bacteria:Q9AKF0	+	149.50	217.38	0.0003	290	! Q9akf0 rickettsia rickettsii. h
sp_bacteria:P94909	+	146.00	206.13	0.0007	529	! P94909 microbacterium ammoniaph
sp_invertebrate:Q9VU43	+	145.00	198.78	0.0011	954	! Q9vu43 drosophila melanogaste
sp_bacteriap:Q92GN2	+	144.50	209.82	0.0008	290	! Q92gn2 rickettsia conorii. lipi
sp_bacteria:Q9AKK5	+	142.50	206.79	0.0012	290	! Q9akk5 rickettsia montana. htrb
sp_human:Q16824	+	142.00	196.02	0.0018	797	! Q16824 homo sapiens (human). pr
sp_human:Q9H4U3	+	142.00	196.01	0.0018	798	! Q9h4u3 homo sapiens (human). dj
sp_bacteria:Q52747	+	141.50	196.84	0.0019	680	! Q52747 ruminococcus flavefacien
sp_bacteriap:Q9CNR4	+	141.00	203.61	0.0017	318	! Q9cnr4 pasteurella multocida. m
sp_bacteria:Q51415	+	140.50	195.27	0.0023	684	! Q51415 pseudomonas aeruginosa.
sp_bacteria:Q9F6E3	+	138.50	199.77	0.0028	320	! Q9f6e3 streptomyces collinus. p
sp_bacteriap:Q9A6A1	+	138.00	196.23	0.0033	424	! Q9a6a1 caulobacter crescentus.
sp_bacteriap:Q9KVD4	+	137.50	198.10	0.0034	325	! Q9kvd4 vibrio cholerae. lipid a
sp_bacteriap:Q9K217	+	137.00	193.86	0.0041	462	! Q9k217 chlamydia pneumoniae (ch
sp_bacteriap:Q9Z983	+	137.00	193.76	0.0041	467	! Q9z983 chlamydia pneumoniae (ch
sp_vertebrate:Q91238	+	133.00	191.27	0.0081	326	! Q91238 pseudopleuronectes ameri
sp_bacteriap:Q9RYP2	+	132.00	188.34	0.0102	376	! Q9ryp2 deinococcus radiodurans.
sp_bacteria:Q53478	+	130.50	184.16	0.0143	456	! Q53478 neisseria gonorrhoeae. c
sp_bacteria:Q93LB1	+	130.00	187.59	0.0141	299	! Q93lb1 bifidobacterium animalis
sp_bacteria:O85018	+	130.00	174.49	0.0201	1122	! O85018 mycobacterium smegmatis
sp_virus:Q9QCQ3	+	128.50	185.39	0.0188	297	! Q9qcq3 pseudorabies virus. t
sp_human:Q9UQ39	+	128.00	173.05	0.0284	956	! Q9uq39 homo sapiens (human).
sp_human:Q9UQ40	+	128.00	170.30	0.0306	1262	! Q9uq40 homo sapiens (human)
sp_human:Q9UHA8	+	128.00	164.37	0.0360	2296	! Q9uha8 homo sapiens (human)
sp_human:Q9UQ35	+	128.00	162.58	0.0378	2752	! Q9uq35 homo sapiens (human)
sp_bacteria:Q9RA26	+	127.50	183.36	0.0231	313	! Q9ra26 vibrio marinus (morit
sp_bacteriap:Q98L49	+	127.50	182.89	0.0234	328	! Q98149 rhizobium loti (mesor
sp_bacteria:O06470	+	127.00	180.05	0.0273	405	! O06470 burkholderia cepacia
sp_invertebrate:O76917	+	126.00	179.22	0.0326	378	! O76917 drosophila melanoga
sp_virus:Q85302	+	125.00	178.44	0.0388	351	! Q85302 orf virus. homologue
sp_plant:Q943U4	+	125.00	164.18	0.0572	1481	! Q943u4 oryza sativa (rice).
sp_bacteria:Q9X2J3	+	124.50	176.77	0.0438	385	! Q9x2j3 ralstonia metallidura
sp_bacteria:Q9AFL5	+	123.50	177.28	0.0503	314	! Q9afl5 shigella flexneri. ac
sp_bacteria:Q9AJU9	+	123.50	177.24	0.0503	315	! Q9aju9 shigella flexneri. ms
sp_plant:Q9SM91	+	123.50	176.49	0.0514	340	! Q9sm91 oryza sativa (rice).
sp_plant:Q9ARV4	+	123.50	175.00	0.0535	395	! Q9arv4 oryza sativa (rice).
sp_bacteria:O05765	+	123.50	170.81	0.0600	603	! O05765 mycobacterium smegmat
sp_plant:Q94DC9	+	123.00	174.63	0.0584	380	! Q94dc9 oryza sativa (rice).
sp_human:O43161	+	123.00	151.69	0.1092	3851	! O43161 homo sapiens (human)
sp_human:Q9UPA5	+	123.00	151.50	0.1098	3926	! Q9upa5 homo sapiens (human)
sp_human:Q9BXI1	+	122.00	163.30	0.0926	1024	! Q9bxi1 homo sapiens (human)
sp_human:Q96EL5	+	122.00	162.41	0.0949	1120	! Q96el5 homo sapiens (human)
sp_human:Q9H7Z4	+	121.50	174.54	0.0736	305	! Q9h7z4 homo sapiens (human).
sp_human:Q96MP7	+	121.50	165.53	0.0941	757	! Q96mp7 homo sapiens (human).
sp_human:Q15206	+	121.50	162.88	0.1011	990	! Q15206 homo sapiens (human).
sp_human:Q05331	+	121.50	160.82	0.1070	1218	! Q05331 homo sapiens (human)
sp_bacteriap:Q9KVD3	+	121.00	173.37	0.0820	318	! Q9kvd3 vibrio cholerae. lipi
sp_rodent:Q99053	+	121.00	166.68	0.0984	625	! Q99053 rattus norvegicus (ra
sp_invertebrate:Q9VUS8	+	120.50	169.28	0.0989	445	! Q9vus8 drosophila melanoga
sp_bacteriap:Q9HUI3	+	120.50	161.28	0.1231	998	! Q9hui3 pseudomonas aeruginos
sp_bacteria:Q9EX24	+	119.50	172.58	0.1053	274	! Q9ex24 streptomyces coelicol
sp_human:Q9H9F1	+	118.50	165.78	0.1477	467	! Q9h9f1 homo sapiens (human).
sp_bacteria:Q9RFF2	+	118.50	163.83	0.1558	569	! Q9rff2 rhodobacter sphaeroid
sp_bacteria:Q9ZI82	+	118.00	168.68	0.1473	323	! Q9zi82 salmonella typhimuriu
sp_bacteria:Q9LCV5	+	118.00	158.68	0.1935	886	! Q9lcv5 rhodospirillum rubrum
sp_invertebrate:Q9VHC0	+	117.50	166.47	0.1689	374	! Q9vhc0 drosophila melanoga
sp_plant:Q9S7V5	+	117.50	162.81	0.1866	541	! Q9s7v5 arabidopsis thaliana
sp_human:Q9UQ36	+	117.50	154.32	0.2353	1275	! Q9uq36 homo sapiens (human)
sp_human:O15038	+	117.50	151.00	0.2576	1783	! O15038 homo sapiens (human)
sp_human:O60382	+	117.50	150.95	0.2579	1791	! O60382 homo sapiens (human)
sp_bacteria:Q9F5M2	+	117.00	169.47	0.1679	256	! Q9f5m2 rhodocyclus gelatinos
sp_invertebrate:P91744	+	117.00	148.18	0.3003	2196	! P91744 loligo bleekeri (b
sp_bacteria:Q9ZA63	+	116.50	168.19	0.1877	270	! Q9za63 neisseria meningitidi
sp_bacteriap:O07148	+	116.50	166.50	0.1965	320	! O07148 mycobacterium leprae.
sp_virus:O57149	+	116.50	155.73	0.2637	949	! O57149 human herpesvirus 6.
sp_bacteriap:Q9RV01	+	116.00	165.78	0.2164	319	! Q9rv01 deinococcus radiodura
sp_virus:Q67851	+	116.00	156.86	0.2760	785	! Q67851 duck hepatitis b viru
sp_invertebrate:Q9UAY0	+	115.00	160.41	0.2919	471	! Q9uay0 caenorhabditis eleg
sp_human:Q01212	+	115.00	152.15	0.3657	1084	! Q01212 homo sapiens (human)
sp_bacteriap:Q92P56	+	114.50	163.73	0.2877	312	! Q92p56 rhizobium meliloti (s
sp_invertebrate:Q9NC84	+	114.50	161.34	0.3070	397	! Q9nc84 strongylocentrotus
sp_bacteriap:Q988B2	+	114.50	161.19	0.3083	403	! Q988b2 rhizobium loti (mesor
sp_bacteriap:Q98DE3	+	114.50	156.52	0.3503	646	! Q98de3 rhizobium loti (mesor
sp_human:O96026	+	114.50	156.32	0.3522	659	! O96026 homo sapiens (human).
sp_plant:Q9M7I5	+	114.00	164.26	0.3060	274	! Q9m7i5 zea mays (maize). ara
sp_human:Q96DG4	+	113.50	153.96	0.4376	718	! Q96dg4 homo sapiens (human).
sp_bacteria:O33629	+	113.00	162.36	0.3755	285	! O33629 saccharopolyspora ery
sp_invertebrate:Q21344	+	113.00	147.95	0.5566	1221	! Q21344 caenorhabditis ele
sp_plant:Q9FQC8	+	112.50	159.02	0.4440	370	! Q9fqc8 zea mays (maize). glu
sp_bacteria:Q9RBG6	+	112.00	158.07	0.4917	377	! Q9rbg6 alcaligenes xylosoxyd

seq_name: sp_bacteriap:Q9JY07

seq_documentation_block:
ID Q9JY07 PRELIMINARY; PRT; 298 AA.
AC Q9JY07;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HTRB/MSBB FAMILY PROTEIN.
GN NMB1801.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

```
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002530; AAF42138.1; -.
DR   TIGR; NMB1801; -.
KW   Complete proteome.
SQ   SEQUENCE   298 AA;  33492 MW;  013437BC4DD6508F CRC64;


alignment_scores:
            Quality: 1469.50              Length:    298
              Ratio:  5.085                Gaps:      1
 Percent Similarity:  96.980   Percent Identity: 94.295

alignment_block:
 US-09-303-518D-571 x Q9JY07   ..

 Align seg 1/1  to: Q9JY07  from: 1  to: 298

      1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

     51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCT 100
        ||||||||||||||||||||||||||||||||||||||||||   |||||||
     17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

    101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

    151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAA 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

    201 CCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCG 250
        |||||||   :::|||||||||||||||||||||||||||||||||   |
     67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

    251 GTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACA 300
        ||||||||||||||||||||||||:::|||||||||||||||||||||||
     84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

    301 ATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAA 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

    351 GGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
        |   ||||||||||||||||||||||||||||||||||||||||||||||||
    117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

    401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTAC 450
        ||||||||||||||||||||||||||||||   ||||||||||||||||
    134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

    451 AAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

    501 GCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAA 550
        ||||||||||||||||||||||||||||||:::|||||||||||||||||
    167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

    551 TCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCAC 600
        ||||||||||||||||:::|||||||||||||||||:::|||||||||||
    184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

    601 GTCCCTTCTCCGCAGGAAGGCGGC...GGCGTGTGGGCGGATTTTTTCGG 647
        ||||||||||||||||||||||   ||||||||:::|||||||||||
    201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

    648 CAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAG 697
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

    698 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGC 747
        ||||||||||||||||||||||||||||||||||   |||||||||
    234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

    748 TTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGC 797
        |||   ||||||||||||||||||||||||||||||||||:::|||||
    251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

    798 CCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTT 847
        ||||||||||||||||||||||||||:::|||||||||||||||||||||
    267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

    848 TTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAAACGCCG 891
        ||||||||||||||||||||||||||||||||||||   |||
    284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: sp_bacteriap:Q9JVV7

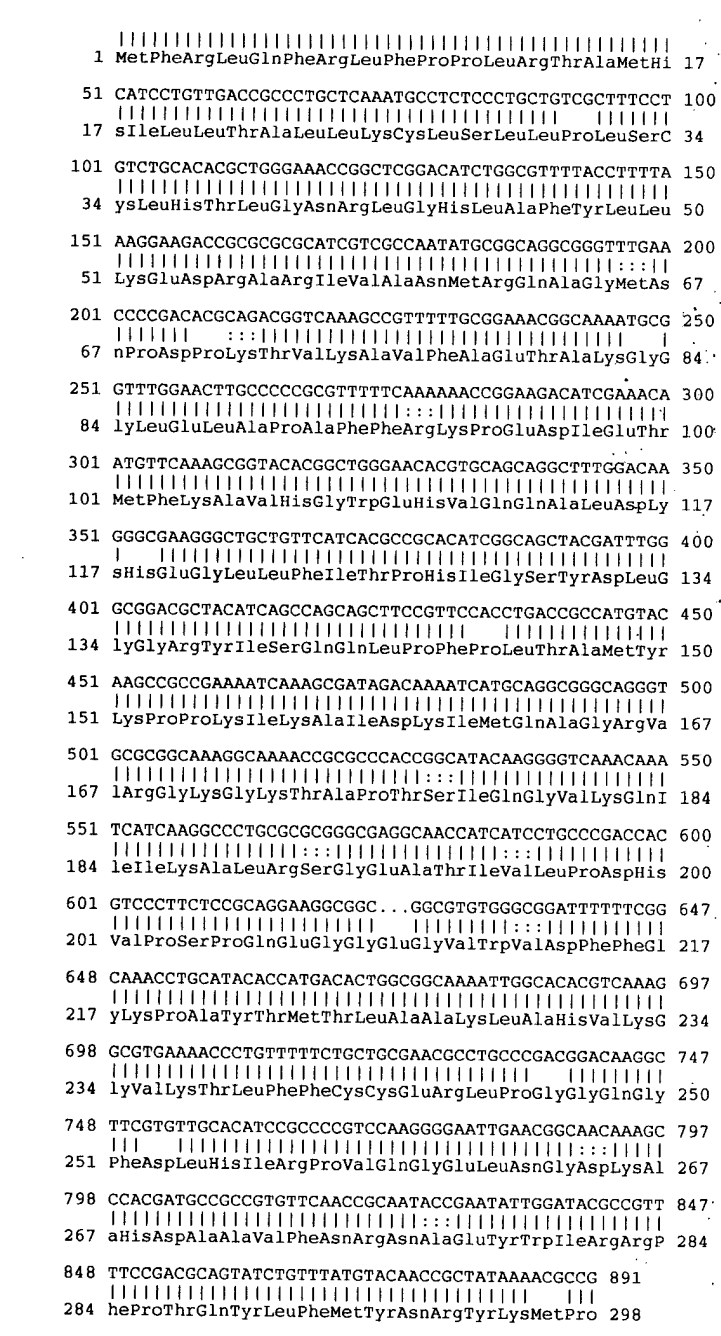
seq_documentation_block:
ID   Q9JVV7      PRELIMINARY;      PRT;   298 AA.
AC   Q9JVV7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE ACYLTRANSFERASE.
GN   NMA0662.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
DR   EMBL; AL162753; CAB83949.1; -.
KW   Transferase; Acyltransferase; Complete proteome.
SQ   SEQUENCE   298 AA;  33510 MW;  2A09946670E283C2 CRC64;


alignment_scores:
            Quality: 1467.50              Length:    298
              Ratio:  5.078                Gaps:      1
 Percent Similarity:  96.980   Percent Identity: 93.960

alignment_block:
 US-09-303-518D-571 x Q9JVV7   ..

 Align seg 1/1  to: Q9JVV7  from: 1  to: 298

      1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
```



seq_name: sp_bacteria:Q9EYM0

```
seq_documentation_block:
ID   Q9EYM0      PRELIMINARY;      PRT;   337 AA.
AC   Q9EYM0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE HTRB PROTEIN.
OS   Acinetobacter sp. BD413.
OC   Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC   Acinetobacter.
OX   NCBI_TaxID=104611;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BD413;
RX   MEDLINE=21318707; PubMed=11425734;
RA   Friedrich A., Hartsch T., Averhoff B.;
RT   "Natural Transformation in Mesophilic and Thermophilic Bacteria:
RT   Identification and Characterization of Novel, Closely Related
RT   Competence Genes in Acinetobacter sp. Strain BD413 and Thermus
RT   thermophilus HB27.";
RL   Appl. Environ. Microbiol. 67:3140-3148(2001).
DR   EMBL; AF320001; AAG34711.1; -.
DR   InterPro; IPR000515; BPD_transp.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
SQ   SEQUENCE   337 AA;  38795 MW;  6E998AFED1CEB140 CRC64;


alignment_scores:
            Quality:  288.00              Length:    278
              Ratio:  1.665                 Gaps:      7
 Percent Similarity:  62.230   Percent Identity: 29.496

alignment_block:
 US-09-303-518D-571 x Q9EYM0   ..

 Align seg 1/1  to: Q9EYM0  from: 1  to: 337

      55 CTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCT 104
         :::|||:::::::|||:::::::   |||:::|||   :::   |||
      31 MetLeuAlaGlyLeuValAsnThrLeuLysLeuThrLysThrSer..... 45

     105 GCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGG 154
          :::|||:::      :::|||                |||||      :
      46 .ArgThrIleGlnLeuAsnLeuLysIleThrLeuProTyrLeuAsnAspG 62

     155 AAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCC 204
         :::::|||   ||||||:::   |||
      62 lnGlnArgGluArgIleIleGlnAsn..................... 70

     205 GACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTT 254
                        |||:::      |||   ::::::      ::
      71 .................AlaIleArgAsnGluLeuThrSerTyr...Me 80

     255 GGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGT 304
         :|||:::         :::   ::::::   :::         :::   :
      80 tGluPhePheSerIleTrpGlySerSerAsnGlnLysAsnIleSerArgI 97

     305 TCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGC 354
         ::      |||:::|||   :::      :::|||||:::   ::::::
      97 leHisHisValGluGlyAlaAspPhePheHisGlnAlaIleAlaGluAsn 113

     355 GAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGG 404
         :::|||:::::::::|||:::|||||||:::|||:::::::::::::   ::
     114 LysGlyIleValLeuIleValProHisPheGlyThrTrpGluIleMetAs 130

     405 ACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGC 454
         :   :::   :::|||               :::|||   :::|||||||
     130 nAlaTrpCysAlaGlnTyrThr......AspMetThrIleLeuTyrLysP 145

     455 CGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGC 504
         ||   |||   |||      |||:::::::::::::   :::|||   |||
     145 roValLysAspLysAspAlaAspArgPheValArgGluAlaArgSerArg 161
```

```
    505 GGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCAT 554
        :::::::|||   :::||||||      :::||||||:::||||||::
    162 GluGlnAlaLysLeuValProThrAspGluSerGlyValArgGlnIlePh 178

    555 CAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCC 604
        :|||||||||:::   |||   :::|||:::|||||||||||||||:::|
    178 eLysAlaLeuLysGlnGlyGlyThrThrValIleLeuProAspHisThrP 195

    605 CTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCT 654
        ||:::            ||||||      :::|||:::||||||   |||
    195 roAsnVal.........GlyGlyGluMetValAspTyrPheGlyIlePro 208

    655 GCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAA 704
            :::   :::|||:::|||||||||   :::         :::||
    209 LeuAlaSerSerAsnLeuSerAlaLysLeuIleGlnLysThrLysAlaLy 225

    705 AACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGT 754
        |:::|||||||:::   :::   |||      :::::::::||||||   :
    225 sAlaLeuPheLeuTyrAlaIleArg...AsnGluAsnHisGlyPheAspM 241

    755 TGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGCCCACGAT 804
        ::   |||:::|||::::::      :::      :::      ||||||
    241 etPheIleGluProMetAspProAlaIleTyrGlnGlyThrAspHisAsp 257

    805 GCCGCC...GTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTCC 851
        ::::::   |||::::::::::      |||      ||||||::::::||
    258 GlyThrLeuValIleHisHisAlaIleGluAspLeuIleArgLysTyrPr 274

    852 GACGCAGTATCTGTTTATGTACAACCGCTATAAA 885
        |   :::|||   :::   |||:::|||:::|||
    274 oAspHisTyrHisTrpSerTyrLysArgPheLys 285
```

seq_name: sp_bacteriap:Q9I7B5

seq_documentation_block:

```
ID   Q9I7B5      PRELIMINARY;      PRT;   295 AA.
AC   Q9I7B5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   PROBABLE 2-OH-LAUROYLTRANSFERASE.
GN   PA0011.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004441; AAG03401.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   295 AA;  33001 MW;  D6D1710EC6AD42FF CRC64;
```

alignment_scores:

```
            Quality:  266.50              Length:    285
              Ratio:  1.514                 Gaps:      7
 Percent Similarity:  61.754   Percent Identity: 26.667
```

alignment_block:

```
 US-09-303-518D-571 x Q9I7B5   ..

 Align seg 1/1  to: Q9I7B5  from: 1  to: 295
```

```
     55 CTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCT 104
        |||:::::::::   |||:::   ::::::||||||         :::::
      8 LeuValValGlyAlaLeuArgLeuPheAlaLeuLeuProTrpArgAlaVa 24

    105 GCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGG 154
        ::::   :::|||      :::|||   |||   :::   |||
     24 lGlnGlyValGlyAlaGlyIleGlyTrpLeuMetTrpLysLeu......P 39

    155 AAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCC 204
          :::|||:::|||   |||:::::::::::   :::         |||
     39 roAsnArgSerArgGluValValArgIleAsnLeuSerLysCysPhePro 55

    205 .........GACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAA 245
              :::|||:::      |||   |||      :::
     56 GluLeuSerGluThrGluLeuGluLysLeuValGlyGlnSerLeuMetAs 72

    246 ATGCGGT.........TTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGG 286
            |||           |||   |||   |||::::::     ||||
     72 pIleGlyArgThrLeuThrGluSerAlaCysAlaTrpIleTrpProProG 89

    287 AAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAG 336
        ||                ::::::   |||:::|||   |||   ::::::
     89 luLysSerLeuArgTyrIleArgGluValGluGlyMetGluValLeuGlu 105

    337 CAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGG 386
        :::||||||   :::|||:::||||||:::   ||||||   |||:::||
    106 GluAlaLeuAlaSerGlyAspGlyLeuValGlyIleThrSerHisLeuGl 122

    387 CAGCTACGAT...TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCC 433
        |:::::::::   |||:::      |||   |||          |||:::
    122 yAsnTrpGluValLeuAsnHisPheTyrCysSerTyrAlaLysProIle. 138

    434 ACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATC 483
                ::::|||:::|||||||||:::||||||:::|||::::::
    139 .........IlePheTyrArgProProLysLeuLysAlaValAspGluLeu 152

    484 ATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCAT 533
        ::::::       ||||||:::       :::::::::|||||||:::
    153 LeuLysLysGlnArgValGlnLeuGlyAsnArgValAlaProSerThrPr 169

    534 ACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGGCGAGGCAACCA 583
        :::|||:::   ::::::||||||   ::::::   |||   ::::::
    169 oGluGlyIleLeuSerValIleLysGluValLysLysGlyGlyCysValG 186

    584 TCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGG 633
          |||       |||       |||:::|||   :::   :::|||::::::
    186 lyIleProAlaAsp......ProGluProAlaArgThrAlaGlyLeuPhe 200

    634 GCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATT 683
        :::   :::::::|||      |||            ::::::   :::||
    201 ValProTyrLeuGlyThrThrAlaLeuIleSerLysPheValProGlnLe 217

    684 GGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGC 733
        |             ::::::   :::||||||   :::   |||||||
    217 uLeuSerArgGlyLysAlaArgGlyValPhePheHisAlaValArgLeuP 234

    734 CCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTG 783
        ||||||||:::|||:::   :::   ::::::   :::   :::::::::
    234 roAspGlySerGlyTyrLysValIleLeuGluAlaAlaProAlaAspMet 250

    784 AACGGCAACAAAGCCCAC...GATGCCGCCGTGTTCAACCGCAATACCGA 830
            :::      :::   :::::::|||:::::::::|||:::
    251 TyrAspLysAspLeuGluValSerValAlaAlaMetSerArgGluLeuAl 267

    831 ATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCT 880
        :::::::|||   :::|||:::||||||:::::      :::|||:
    267 aLysTyrValArgAlaTyrProSerGlnTyrMetTrpSerMetLysArgP 284

    881 ATAAA 885
        ::|||
    284 heLys 285
```

```
seq_name: sp_bacteriap:Q9PDN1

seq_documentation_block:
ID   Q9PDN1      PRELIMINARY;      PRT;   316 AA.
AC   Q9PDN1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE.
GN   XF1348.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE003967; AAF84157.1; -.
KW   Complete proteome.
SQ   SEQUENCE   316 AA;  36286 MW;  8450EED03A405D3A CRC64;


alignment_scores:
            Quality:  258.00              Length:    283
              Ratio:  1.518                 Gaps:      9
 Percent Similarity:  60.071   Percent Identity: 27.915

alignment_block:
 US-09-303-518D-571 x Q9PDN1   ..

 Align seg 1/1  to: Q9PDN1  from: 1  to: 316

     103 CTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAA 152
         |||:::   :::::::::   :::|||||||:::   :::   :::|||||
      10 LeuTyrArgValAlaSerLeuValGlyHisIleProTrpProPheLeuLy 26

     153 GGAAGACCGCGCGCGCATCGTCGCC.................AATATGC 184
         |:::      :::   :::::::|||                |||   |
      26 sArg...LeuSerAspLeuIleAlaTyrCysTrpLeuCysLeuAsnAlaA 42

     185 GGCAGGCG.................................GGTTTG 198
         ||::::::                                   |||
      42 rgGluSerArgValThrGlnArgAsnLeuGluLeuThrTyrProGluLeu 58

     199 AACCCCGAC......ACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGC 242
         :::|||:::      |||:::            ::::::   :::|||::
      59 SerProGlnGluHisThrArgLeuHisHisGlnIleIleTyrSerThrVa 75

     243 AAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACA 292
         ::::   ::::::|||:::         :::                 :::
      75 lArgGlnAlaPheGluMetLeuHisIleTrpThrHisProGluThrGluA 92

     293 TCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCT 342
           :::   ::::::      ||||||   :::      :::   |||
      92 snLeuAlaArgLeuArgGluTyrHisGlyValAspLeuTyrSerAlaAla 108

     343 TTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTA 392
         |||   ::::::::::|||:::::::::::::::::|||||:::|||:::::
     109 LeuAlaArgAsnArgGlyMetIleIleValValProHisPheGlyAsnTr 125

     393 CGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCG 442
         :::|||      :::|||:::|||:::      |||      ||||||
     125 pGluLeuLeuLysGlnTyrLeuSerAspLeuAlaPro......LeuThrL 140

     443 CCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCG 492
           ||||||:::|||   :::      ::::::|||   :::::::|||
     140 euMetTyrArgProAlaAsnSerAlaValLeuAspGlyPheLeuGln... 155

     493 GGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGG.. 540
           ||||||||||||   :::::::::      :::      :::|||
     156 ...ArgValArgGlyGlyAsnAsnValHisGlnValArgAlaGluGlyMe 171

     541 ....GTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCA 586
             |||:::::::::::::|||:::|||:::   |||   :::      |
     171 tAlaValArgHisLeuPheLysValLeuLysAsnGlyGlyThrIleGlyI 188

     587 TCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCG 636
         ||||||||||||:::      :::|||:::   |||   ||||||:::|||
     188 leLeuProAspGln......GlnProLysIleGlyGluGlyValPheAla 202

     637 GATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGC 686
           |||||||||      |||   ||||||||||||:::   :::|||||
     203 ProPhePheGlyIleGluAlaLeuThrMetThrLeuValGlnArgLeuAl 219

     687 ACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCG 736
         |:::      ::::::   :::|||:::      |||||||||:::
     219 aGluArgThrSerAlaThrValLeuTyrAlaTrpCysGluArgIleSerP 236

     737 ACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTG... 783
                 |||:::||||||:::::   ::::::      :::
     236 roHisLeuGluPheAlaLeuHisMetGluHisAlaAspProAlaValAla 252

     784 AACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAATACCGAATA 833
         :::   :::            |||:::::::::|||   :::   |||
     253 AspProAspProLeuIleAlaAlaThrAlaLeuAsnAlaGlyIleGluAr 269

     834 TTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTAT 882
               ||||||   ||||||||||||   :::   |||:::||||||
     269 gIleAlaArgArgAspProThrGlnTyrGlnTrpThrTyrLysArgTyr 285

seq_name: sp_bacteria:O33437

seq_documentation_block:
ID   O33437      PRELIMINARY;      PRT;   253 AA.
AC   O33437;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HTRB PROTEIN (FRAGMENT).
GN   HTRB.
OS   Pseudomonas fluorescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=294;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WCS365;
```

```
RX   MEDLINE=98340543; PubMed=9675892;
RA   Dekkers L.C., van der Bij A.J., Mulders H.M., Phoelich C.C.,
RA   Wentwoord A.R., Glandorf D.C.M., Wijffelman C.A., Lugtenberg B.J.J.;
RT   "Role of the O-antigen of lipopolysaccharide, and possible roles of
RT   growth rate and NADH:ubiquinone oxidoreductase (nuo) in competitive
RT   tomato root-tip colonization by Pseudomonas fluorescens WCS365.";
RL   Mol. Plant Microbe Interact. 11:763-771(1998).
DR   EMBL; Y14568; CAA74897.1; -.
FT   NON_TER     253    253
SQ   SEQUENCE   253 AA;  27634 MW;  F12E003B10C8223D CRC64;


alignment_scores:
            Quality:  226.00              Length:    248
              Ratio:  1.477                 Gaps:      8
 Percent Similarity:  61.694   Percent Identity: 29.435

alignment_block:
 US-09-303-518D-571 x O33437   ..

 Align seg 1/1  to: O33437  from: 1  to: 253

      55 CTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCT 104
         ||||||::::::   |||:::   ::::::||||||         :::::
       8 LeuLeuValGlyAlaLeuArgLeuPheAlaLeuLeuProTrpArgAlaVa 24

     105 GCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGG 154
         ::::::::::|||:::   :::|||   :::   :::   |||
      24 lGlnAlaValGlySerAlaIleGlyTrpPheMetTrpLysLeu......P 39

     155 AAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCC 204
           :::|||:::|||   |||:::::::::::   |||         |||
      39 roAsnArgSerArgAspValValArgIleAsnLeuAlaLysCysPhePro 55

     205 GAC.................ACGCAGACGGTCAAAGCCGTTTTTGCGGA 236
         :::                 :::   |||         :::   ::
      56 GluMetAspProAlaGluArgGluArgLeuValGlyGlnSerLeuLysAs 72

     237 AACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGG 286
         :   :::|||         |||   |||   |||::::::         :
      72 pIleGlyLysSerLeuThrGluSerAlaCysAlaTrpIleTrpProAlaG 89

     287 AAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAG 336
         ::          :::   :::   |||:::|||   |||   ::::::
      89 lnArgSerIleAspLeuValArgGluValGluGlyLeuGluValLeuLys 105

     337 CAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGG 386
         :::||||||   :::|||:::|||::::::   ||||||   |||:::||
     106 GluAlaLeuAlaSerGlyLysGlyValValGlyIleThrSerHisLeuGl 122

     387 CAGCTACGAT...TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCC 433
         |:::::::::   |||:::      |||   ||||||      |||:::
     122 yAsnTrpGluValLeuAsnHisPheTyrCysSerGlnCysLysProIle. 138

     434 ACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATC 483
                   :::|||:::|||||||||:::||||||:::|||::::::
     139 ........IlePheTyrArgProProLysLeuLysAlaValAspGluLeu 152

     484 ATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCAT 533
         ::::::      ||||||:::      :::|||:::|||   :::
     153 LeuArgLysGlnArgValGlnLeuGlyAsnLysValAlaAlaSerThrLy 169

     534 ACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCA 583
          :::|||:::   :::::::|||||   :::|||   |||       :::
     169 sGluGlyIleLeuSerValIleLysGluValArgLysGlyGlyGlnValG 186

     584 TCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGG 633
           |||       |||       |||:::|||   |||::::::|||::::::
     186 lyIleProAlaAsp......ProGluProAlaGluSerAlaGlyIlePhe 200

     634 GCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAA.. 681
         :::   ||||||:::      |||   |||      ::::::   :::
     201 ValProPhePheAlaThrGlnAlaLeuThrSerLysPheValProAsnMe 217

     682 .TTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCC 730
          ||||||      |||:::|||      :::|||:::   :::   ||||
     217 tLeuAlaGlyGlyLysAlaVal...GlyValPheLeuHisAlaLeuArgL 233

     731 TGCCCGACGGACAAGGCTTC......GTGTTGCACATCCGCCCC 768
         ||||||||||||:::|||:::      :::|||:::      |||
     233 euProAspGlySerGlyTyrLysValIleLeuGluAlaAlaPro 247

seq_name: sp_bacteriap:Q9HYZ8

seq_documentation_block:
ID   Q9HYZ8      PRELIMINARY;      PRT;   312 AA.
AC   Q9HYZ8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   PROBABLE LAUROYL ACYLTRANSFERASE.
GN   PA3242.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004747; AAG06630.1; -.
KW   Transferase; Acyltransferase; Complete proteome.
SQ   SEQUENCE   312 AA;  36063 MW;  0EDE771D75F3C186 CRC64;


alignment_scores:
            Quality:  222.50              Length:    311
              Ratio:  1.236                 Gaps:     11
 Percent Similarity:  57.878   Percent Identity: 26.045

alignment_block:
 US-09-303-518D-571 x Q9HYZ8   ..

 Align seg 1/1  to: Q9HYZ8  from: 1  to: 312

      33 TTTGCGAACCGCCATGCACATCCTGTTGACCGCCCT..... 68
         |||::::::         ||||||         |||
       6 PheSerSerAlaPheLeuHisProArgTyrTrpProLeuTrpPheGlyLe 22

      69 ..................GCTCAAATGCCTCTCCCTGCTGTCGCTTTCC 99
                           :::|||:::|||   |||:::::   :::|
      22 uGlyLeuLeuTrpLeuValValGlnLeuProTyrProValLeuLeuMetL 39

     100 TGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTT 149
         ||                  :::|||      |||:::|||   :::
      39 eu..................GlyArgGlyLeuGlyAlaLeuMetTyr 48

     150 AAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGA 199
               |||      :::   ||||||:::   ||||||   ::::::||
      49 ......ArgLeuValGlySerArgArgGluIleAlaAlaArgAsnLeuGl 63

     200 A.........................CCCCGACACGCAGACGGTCAAA 222
         |                        |||::: :::   :::|||
      63 uLeuCysPheProGluLysSerProAlaGluArgGlu.ArgLeuLeuLys 79

     223 GCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTT 272
               ||||||:::::::::   :::   :::|||:::|||   :::::
```

```
 80 GluAsnPheAlaSerSerGlyIleAlaPhePheGluMetAlaMetSerTr 96

273 TTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCT 322
    ::::      |||:::         :::   :::       ::::::|||
 96 pTrp...TrpProLysAlaArgLeuAlaArgLeuAlaHisIleGluGlyL 112

323 GGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATC 372
      ||||||:::::::::|||     :::||||||||||:::::::::::
112 euGluHisLeuArgGluAlaGlnAlaGlnGlyGluGlyValIleLeuMet 128

373 ACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCA 422
    :::   |||:::   :::   ::::::|||:::      ::::::|||
129 AlaLeuHisPheThrThrLeuGluIleGlyAlaAlaLeuLeuGlyGln.. 144

423 GCTTCCGTTCCAC...CTGACCGCCATGTACAAGCCGCCGAAAATCAAAG 469
              |||   :::   :::|||||:::              :
145 .......ValHisThrIleAspGlyMetTyrArgGluHisAspAsnProV 159

470 CGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACC 519
    :::::|||   :::::::::   ||||||   |||   :::      :::
159 alPheAspTyrValGlnArgArgGlyArgGluArgHisAsnLeuAspAla 175

520 GCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGC 569
    :::           :::   |||:::::::::::::|||:::||||||::
176 ThrAlaIleGluArgGluAspValArgSerMetLeuLysValLeuArgGl 192

570 GGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAG 619
    :|||:::|||           ||||||          |||:::
192 yGlyArgAlaIleTrpTyrAlaProAsp.............GlnAspT 204

620 GCGGC.........GGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATAC 660
      |||         ::::::::::::   :::||||||   ||||||
204 yrGlyAlaLysGlnSerLeuPheValProLeuPheGlyIleProAlaAla 220

661 ACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCT 710
    |||:::|||   :::::::|||:::|||::::::      ::::::::||
221 ThrValThrAlaThrThrLysPheAlaArgLeuGlyArgAlaArgValLe 237

711 GTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACA 760
    |   |||      :::||||||   ||||||:::|||:::   |||   |
237 uProPheThrGlnSerArgLeuAlaAspGlySerGlyTyrArgLeuThrI 254

761 TCCGCCCCGTCCAAGGGGAATTGAACGGC...AACAAAGCCCACGATGCC 807
    ||:::|||         ::::::   |||   ::::::      |||:::
254 leHisProProLeuGluAspPheProGlyGluSerGluGluAlaAspCys 270

808 GCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTCCGACGCA 857
          :::|||:::   :::|||      :::|||:::   |||   ||
271 LeuArgIleAsnGlnTrpValGluArgCysValArgGlnGlnProGluGl 287

858 GTATCTGTTTATGTACAACCGCTATAAAACG 888
    |||||||:::   ::::::|||:::||||||
287 nTyrLeuTrpAlaHisArgArgPheLysThr 297

seq_name: sp_bacteriap:Q9JYV2

seq_documentation_block:
ID   Q9JYV2      PRELIMINARY;      PRT;   289 AA.
AC   Q9JYV2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HTRB/MSBB FAMILY PROTEIN.
GN   NMB1418.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002491; AAF41779.1; -.
DR   TIGR; NMB1418; -.
KW   Complete proteome.
SQ   SEQUENCE   289 AA;  33843 MW;  3BF100F050576512 CRC64;


alignment_scores:
            Quality:  204.00              Length:    289
              Ratio:  1.207                 Gaps:     12
 Percent Similarity:  58.478   Percent Identity: 27.336

alignment_block:
 US-09-303-518D-571 x Q9JYV2   ..

 Align seg 1/1  to: Q9JYV2  from: 1  to: 289

 67 CTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGG 116
    :::|||      |||:::::::|||  ::::::  ||||||  :::::
  7 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysIleAl 23

117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
    ::::      |||  ||||||  |||||||||  ::  |||  |
 23 aAspLeuThrGlyLeuLeuAla...TyrLeuLeuValLysProArgArgA 39

167 GCATC...........GTCGCCAATATGCGGCAGGCGGGTTTGAACCCC 204
    |||||           :::|||:::     :::         :::
 39 rgIleGlyGluIleAsnLeuAlaLysCysPheSerGluTrpSerGluGlu 55

205 GACACGCAGACGGTC...AAAGCCGTTTTTGCGGAAACGGCAAAATGCGG 251
          :::||||||  |||     |||  :::  ||||||
 56 LysArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMe 72

252 TTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAA 301
     ||||||  :::      :::::                 ::::::::::
 72 tLeuGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerL 89

302 TGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAG 351
    ::  :::      :::::  ::::::::::::::|||||
 89 euValArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAla 104

352 GGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGG 401
    ||||||  ::::::::::::  ||||||:::  ::::::::::::::
105 GlyGluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAl 121

402 CGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACA 451
    :       ::::::|||:::::::|||      |||  :::||||||:
121 aValTyrAlaLeuAsnGlnAspIlePro......LeuIleSerMetTyrS 136

452 AGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTG 501
    ::      |||  |||  :::|||:::  :::      ||||||
136 erHisGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArgAsn 152

502 CGCGGCAAA.............GGCAAAACCGCGCCCACCGGCATACA 536
    |||                  |||:::|||             ::
153 ArgTyrHisAsnValPheLeuIleGlyArgThr.............Gl 164

537 AGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCA 586
    :|||::::::  ::::::|||  :::|||  :::::::|||  ::::
164 uGlyLeuArgAlaLeuValLysGlnPheArgLysSerSerAlaProPheL 181

587 TC...CTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGC..... 624
    ::  |||||||||             |||:::  |||
181 euTyrLeuProAsp.............GlnAspPheGlyArgAsnAsp 192
```

```
     625 GGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGC 674
         :::|||::::::|||||||||||||      :::   |||:::|||
     193 SerValPheValAspPhePheGlyIleGlnThrAlaThrIleThrGlyLe 209

     675 GGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCG 724
         ::::::::::|||   :::   :::::::|||::::::
     209 uSerArgIleAlaAlaLeuAlaAsnAlaLysValIleProAlaIleProV 226

     725 AACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAA 774
           |||      |||:::         :::|||||||:::   |||:::
     226 alArgGluAlaAspAsnThr...ValThrLeuHisPheTyrProAlaTrp 241

     775 GGGGAATTGAACGGCAACAAAGCCCAC...GATGCCGCCGTGTTCAACCG 821
           ::::::   |||:::   |||      ||||||      :::|||||
     242 LysSerPheProGlyGluAspAlaLysAlaAspAlaGlnArgMetAsnAr 258

     822 CAATACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGT 871
         |      |||      :::|||:::   |||   ||||||:::::::::::
     258 gPheIleGluAspArgValArgGluHisProGluGlnTyrPheTrpLeuH 275

     872 ACAACCGCTATAAAACG 888
         :::::|||:::||||||
     275 isLysArgPheLysThr 280

seq_name: sp_bacteriap:Q9JTU4

seq_documentation_block:
ID   Q9JTU4      PRELIMINARY;      PRT;   289 AA.
AC   Q9JTU4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE ACETYLTRANSFERASE.
GN   NMA1630.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
DR   EMBL; AL162756; CAB84858.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   289 AA;  33867 MW;  4FAE453B5A632C1D CRC64;


alignment_scores:
            Quality:  202.00            Length:    288
              Ratio:  1.195               Gaps:     10
 Percent Similarity:  58.681   Percent Identity: 26.389

alignment_block:
 US-09-303-518D-571 x Q9JTU4   ..

 Align seg 1/1  to: Q9JTU4  from: 1  to: 289

      67 CTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGG 116
         :::|||      |||:::::::|||   ::::::   ||||||   |||::
       7 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysLeuAl 23

     117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
         ::::      |||   ||||||:::   |||:::|||      |||
      23 aAspLeuThrGlyLeuLeuAlaTyrLeuLeuValLysProArgArgArgI 40

     167 GCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACG 216
                   |||:::   ::::::      :::   |||   :::
      40 leGlyGluIleAsnLeuAlaLysCysProGluTrpAspGlyLysLys 56

     217 GTCAAAGCCGTT.....TTTGCGGAAACGGCAAAATGCGGTTT 254
         |||:::|||           |||   :::   ||||||      ||
      57 ArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMetLe 73

     255 GGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGT 304
         ||||   :::      ::::::                :::::::::::
      73 uGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerLeuV 90

     305 TCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGC 354
           :::      ::::::   ::::::::::::::::||||||      |||
      90 alArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAlaGly 105

     355 GAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGG 404
         |||   ::::::::::::   |||||||:::   ::::::::::::::
     106 GluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAlaVa 122

     405 ACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGC 454
               ::::::|||::::::|||      |||   :::|||||||:::
     122 lTyrAlaLeuAsnGlnAspValPro......LeuIleSerMetTyrSerH 137

     455 CGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGC 504
           |||   |||   :::|||:::   :::      ||||||   |||
     137 isGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArgAsnArg 153

     505 GGCAAA.............GGCAAAACCGCGCCCACCGGCATACAAGG 539
                            |||:::|||               :::||
     154 TyrHisAsnValPheLeuIleGlyArgThr.............GluGl 165

     540 GGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATC. 588
         |::::::   ::::::|||   :::|||   ::::::|||   ::::::
     165 yLeuArgAlaLeuValLysGlnPheArgLysSerSerAlaProPheLeuT 182

     589 ..CTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGC.........GGC 627
           |||||||||               |||:::   |||       :::
     182 yrLeuProAsp.............GlnAspPheGlyArgAsnAspSer 193

     628 GTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGC 677
         |||::::::|||||||||||||      :::   |||:::|||      ::
     194 ValPheValAspPhePheGlyIleArgThrAlaThrIleThrGlyLeuSe 210

     678 AAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAAC 727
         :::::::|||   :::   ::::::|||:::::::                |
     210 rArgIleAlaAlaLeuAlaAsnAlaLysValIleProAlaIleProValA 227

     728 GCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGG 777
         ||      |||:::         :::||||||:::   |||:::
     227 rgGluAlaAspAsnThr...ValThrLeuHisPheTyrProAlaTrpGlu 242

     778 GAATTGAACGGCAACAAAGCCCAC...GATGCCGCCGTGTTCAACCGCAA 824
         ::::::   ::::::   |||:::   ||||||      :::||||||
     243 SerPheProSerGluAspAlaGlnAlaAspAlaGlnArgMetAsnArgPh 259

     825 TACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACA 874
             |||      :::|||:::   |||   ||||||:::::::::::::
     259 eIleGluGluArgValArgGluHisProGluGlnTyrPheTrpLeuHisL 276

     875 ACCGCTATAAAACG 888
         ::|||:::||||||
     276 ysArgPheLysThr 280

seq_name: sp_bacteriap:Q9PH42

seq_documentation_block:
ID   Q9PH42      PRELIMINARY;      PRT;   279 AA.
AC   Q9PH42;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE.
```

```
GN   XF0104.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE003864; AAF82917.1; -.
KW   Complete proteome.
SQ   SEQUENCE   279 AA;  32072 MW;  450EC409876A7791 CRC64;


alignment_scores:
            Quality:  192.00              Length:    270
              Ratio:  1.255                 Gaps:     10
 Percent Similarity:  56.667   Percent Identity: 27.407

alignment_block:
 US-09-303-518D-571 x Q9PH42   ..

 Align seg 1/1  to: Q9PH42  from: 1  to: 279

     112 CTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCG 161
         ||||||:::   :::|||   :::|||:::   |||:::          ||
      15 LeuGlyArgValIleGlyAlaMetAlaLeuArgLeuValGlyThrArgAr 31

     162 CGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGC 211
         |:::         :::|||:::|||   :::       :::   |||
      31 gCysAlaAlaGluValAsnLeuArgLeuCysPheProThrHisAspGluA 48

     212 AGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGC.........GGT 252
                |||   ::::::   |||:::      :::         |||
      48 laTrpArgLysLysLeuLeuHisGluAsnPheAspAlaLeuGlyValGly 64

     253 TTG...GAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAAC 299
         |||   |||:::|||   |||::::::            :::|||:::::
      65 LeuPheGluPheAlaArgAlaTrpTrp.........GlySerIleAspVa 78

     300 AATGTTCAAAGCGGTACAC......GGCTGGGAACACGTGCAGCAGGCTT 343
         ::::      ::::::|||      |||   ||||||::::::|||
      78 lIleArgAlaGlyIleHisValGluGlyLeuGluHisLeuHisGlnLeuG 95

     344 TGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTAC 393
                ::::|||:::|||:::|||:::::::::   |||:::   :::
      95 lnArgGluGlyArgGlyValLeuLeuValSerGlyHisPheMetThrLeu 111

     394 GATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGC 443
         ::::::   ||||||   :::   :::   :::|||      |||:::::
     112 GluIleCysGlyArgLeuLeuCysAspTyrValPro......LeuAlaGl 126

     444 CATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGG 493
         :|||||:::      :::      :::::::::      ::::::   |
     126 yMetTyrArgArgHisArgAsnProValLeuGluTrpAlaIleLysArgG 143

     494 GCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTC 543
         |||||:::|||          |||         :::   :::   :::
     143 lyArgLeuArg......TyrAlaThrGlnMetPheAlaAsnGluGluLeu 157

     544 AAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCC 593
         :::      ||||||   |||:::   |||                   ||
     158 ArgAlaThrIleLysHisLeuLysArgGlyGlyPheLeuTrpTyrAlaPr 174

     594 CGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTT 643
         ||||:::             :::|||         |||:::|||   ||||
     174 oAspGlnAspMetArgGlyLysGlu......ThrValPheAlaProPheP 189

     644 TCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTC 693
         |||||   ||||||   |||:::|||   :::   :::||||||:::::
     189 heGlyMetProAlaAlaThrIleThrAlaThrHisGlnLeuAlaArgLeu 205

     694 AAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCC........ 735
           |||                     |||      :::|||
     206 ThrGly.....................CysAlaValValProTyrPheHi 215

     736 .......GACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGG 778
                 :::|||   ::::::||||||   |||   |||      ::::
     215 sArgArgGlnGlyGlyAsnTyrValLeuLysIleAlaProProLeuAlaA 232

     779 AATTG...AACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAAT 825
         :::::   :::   :::   |||   |||:::|||   :::|||:::
     232 spPheProSerGluAspLeuAlaAlaAspThrAlaArgIleAsnGlnVal 248

     826 ACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAA 875
           |||      |||   :::   |||   |||||||||:::::::::::
     249 IleGluThrMetIleAspGluAlaProAspGlnTyrLeuTrpIleHisAr 265

     876 CCGCTATAAA 885
         :|||:::|||
     265 gArgPheLys 268

seq_name: sp_bacteria:Q9API8

seq_documentation_block:
ID   Q9API8      PRELIMINARY;      PRT;   292 AA.
AC   Q9API8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE LAYROYL ACYLTRANSFERASE.
OS   Methylobacterium extorquens.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Methylobacterium group; Methylobacterium.
OX   NCBI_TaxID=408;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AM1;
RX   MEDLINE=21142531; PubMed=11208803;
RA   Korotkova N., Lidstrom M.E.;
RT   "Connection between Poly-beta-Hydroxybutyrate Biosynthesis and Growth
RT   on C1 and C2 Compounds in the Methylotroph Methylobacterium extorquens
RT   AM1.";
RL   J. Bacteriol. 183:1038-1046(2001).
DR   EMBL; AF287907; AAK11538.1; -.
KW   Transferase; Acyltransferase.
SQ   SEQUENCE   292 AA;  32214 MW;  02AD260FA71E7C11 CRC64;
```

```
alignment_scores:
            Quality:  178.00              Length:    293
              Ratio:  1.187                 Gaps:     12
 Percent Similarity:  51.195   Percent Identity: 24.915

alignment_block:
 US-09-303-518D-571 x Q9API8   ..

 Align seg 1/1  to: Q9API8  from: 1  to: 292

     102 TCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAA 151
           ||||||   :::|||:::   :::|||       |||      ||||||
       8 SerAlaAlaGlyGlyArgLeuValArgLeuIleGly......ProPheLe 22

     152 AGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAAC 201
               |||:::   :::   |||:::|||   |||      :::
      22 uProSerHisArgThr.AlaMetAspAsnLeuArgAlaAlaTyrProLys 38

     202 CCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACG..... 240
           |||   :::   :::::::|||:::      |||:::
      39 GluAspGluArgArgLeuArgAlaIleAlaLeuGluAlaTrpAspAsnLe 55

     241 .GCAAAATGCGGTTTGGAACTTGCC.........CCCGCGTTTTTCAAAA 280
          ::::::   |||   |||   |||              |||
      55 uGlyArgThrGlyAlaGluTyrAlaHisLeuAspThrIlePheAspTyrA 72

     281 AACCGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACAC 330
           ||||||::::::   |||:::         |||:::|||   |||:::
      72 spProGluAsnLeuAlaThrGlnArgMetGluValArgGlyLeuGluGln 88

     331 GTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCA 380
                       |||   |||:::      |||::::::::::   ||
      89 PhePheAlaIleArgAspAspGlyGlnProGlyLeuIlePheSerAlaHi 105

     381 CATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGT 430
          |:::::::::::::::|||   :::          ::::::::::::   :
     105 sLeuAlaAsnTrpGluLeuProAla...IleCysAlaGluLysPheGlyL 121

     431 TCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAA 480
          :::::   ||||||:::::::::|||||||:::      |||      :::
     121 euGluThrThrAlaValPheArgProProAsnAsnProAlaAlaAlaGln 137

     481 ATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGG 530
          :::::::|||      |||   :::      |||      |||   :::||
     138 LeuValGlnGluValArgArgLysThrMetGlyGlyLeuAlaAlaSerGl 154

     531 C.................ATACAAGGGGTCAAACAAATCATCAAGGCCC 562
          |                :::|||||||||   :::
     154 yProGlyAlaValPheAlaMetGlnGlyValValGlu............. 166

     563 TGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCG 612
            |||:::|||:::         ::::::   :::|||
     167 ..ArgGlyGlyHisLeuGlyGlnLeuIleAspGlnHisPheThrArg... 181

     613 CAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACAC 662
                   ||||||   :::   |||||||||:::|||:::   ::
     182 ............GlyValValValThrPhePheGlyArgProValLeuVa 194

     663 CATGACACTGGCGGCAAAATTGGCACAC..............GTCAAAG 697
          :    |||   :::|||||||||:::               |||   |
     194 lAsnProLeuLeuGlyLysLeuAlaArgHisHisAspCysProValHisG 211

     698 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGC 747
          ||:::::::::::                          |||
     211 lyAlaArgValVal.................ArgLysGlyGlyGlyArg 221

     748 TTCGTGTTGCACATCCGCCCC............................. 768
          |||:::|||::::::   |||
     222 PheLeuLeuGluLeuThrProProLeuAspLeuProArgGlyProAspGl 238

     769 .........GTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCG 808
                 |||||||||   ::::::

     238 yLeuIleGluValGlnGlyAlaMetGln..................... 247

     809 CCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTCCGACGCAG 858
           :::::::::|||      |||   |||:::|||:::   |||   |||
     248 ..AlaMetThrArgValIleGluGlyTrpValArgGluHisProGlyGln 263

     859 TATCTGTTTATGTACAACCGCTATAAA 885
          :::|||:::|||:::::::|||::::::
     264 TrpLeuTrpMetHisArgArgTrpArg 272

seq_name: sp_bacteria:Q9L7Q3

seq_documentation_block:
ID   Q9L7Q3      PRELIMINARY;      PRT;   317 AA.
AC   Q9L7Q3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HTRB.
GN   HTRB.
OS   Haemophilus ducreyi.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=730;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Munson R.S. Jr., Bozue J.A.;
RT   "Identification of the htrB gene in Haemophilus ducreyi.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF221008; AAF34642.1; -.
SQ   SEQUENCE   317 AA;  36521 MW;  8359FE1D5945612B CRC64;


alignment_scores:
            Quality:  176.00              Length:    286
              Ratio:  1.067                 Gaps:     12
 Percent Similarity:  57.692   Percent Identity: 26.224

alignment_block:
 US-09-303-518D-571 x Q9L7Q3   ..

 Align seg 1/1  to: Q9L7Q3  from: 1  to: 317

      64 GCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCT 113
         |||::::::|||   :::      |||             |||      ::
      32 AlaIlePheLysLeuIleLeuCysLeuProTyrProIleLeuValLysIl 48

     114 GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCG 163
         :|||:::   |||:::   |||   |||      |||:::      :::
      48 eGlyArgSerLeuAlaLysLeu...PheSerThrLeuSerPheGlyLysA 64

     164 CGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTG...AACCCCGAC... 207
           ||||||::::::      |||:::   :::|||      ||||||
      64 rgArgIleAlaValAlaArgArgAsnLeuAlaLeuCysPheProAspTyr 80

     208 ...ACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAA........ 246
          ::::::   :::::::|||::::::   ::::::::::::
      81 SerAlaGluGlnIleGluAlaIleLeuIleLysAsnValGluSerValGl 97

     247 ....TGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACA 292
             :::   :::|||   :::   |||:::|||   :::   :::   |
      97 yMetAlaIleIleGluThrGlyMetAlaTrpPheTrpSerAspLysArgI 114

     293 TCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCT 342
          ||        ::::::   :::::::|||   |||||||:::||||||
     114 le...LeuLysTrpSerAsnIleGluGlyLeuGluHisLeuGlnGln... 128

     343 TTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTA 392
               ||||||   |||:::|||||||:::      |||:::   :::
     129 ...ProLysGlyThrGlyIleLeuPheValGlyValHisPheLeuThrLe 144

     393 CGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCG 442
          :::||||||:::|||   ::::::      |||      :::      :
```

```
     144 uGluLeuGlyAlaArgIleValGly......LeuHisHisGlnGlyIleG 159

     443 CCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCG 492
         :::::|||:::|||                 ::::::   |||:::
     159 lyValTyrArgProAsnAspAsnProLeuLeuAsnTrpIleGlnPheLys 175

     493 GGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGT 542
         ||||||:::|||      ::::::   |||          :::   ::
     176 GlyArgIleArg......SerAsnLysAlaMetLeuAspArgLysAspLe 190

     543 CAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGC 592
         ::::   :::|||:::|||||||||   ||||||:::             |
     190 uArgGlyMetIleArgAlaLeuArgAspGlyGluThrIleTrpTyrAlaP 207

     593 CCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTT 642
         ||||||||      :::   |||      ::::::|||::::::   |||
     207 roAspHisAspTyrGlyGlnGln......AsnSerValPheValProPhe 221

     643 TTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGT 692
         |||:::                  |||   ::::::   :::
     222 PheAlaValGluGlnAlaCysThrThrAlaGlySerHisMetLeuLeuAr 238

     693 CAAAGGCGTGAAAACCCTGTTT......TTCTGCTGCGAACGCCTGCCCG 736
          ::::::   :::::::::         |||         |||      |
     238 gSerAlaProAsnSerIleValValProPheThrProIleArgHisTyrA 255

     737 ACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAAC 786
         ||   :::|||:::::::::   |||   |||:::      ::::::|||
     255 spPheSerGlyTyrThrValLysIleSerProAlaVal...AspPheAsn 270

     787 GGC......AACAAAGCCCACGATGCCGCCGTGTTCAACCGCAATACCGA 830
               ||||||   :::   |||   ::::::|||:::   :::||
     271 AspCysGlnAsnLysIleGluThrAlaArgLeuMetAsnLysValValGl 287

     831 ATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCT 880
         |      |||   :::      |||||||||:::::::::::::::|||:
     287 uThrGluIleLeuLysAlaGlnThrGlnTyrMetTrpLeuHisArgArgP 304

     881 ATAAAACG 888
         ::||||||
     304 heLysThr 306

seq_name: sp_bacteria:Q9L7U9

seq_documentation_block:
ID   Q9L7U9      PRELIMINARY;      PRT;   318 AA.
AC   Q9L7U9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   LIPID A ACYLTRANSFERASE.
GN   MSBB.
OS   Haemophilus ducreyi.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=730;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Munson R.S. Jr., Bozue J.A.;
RT   "Identification of the msbB gene in Haemophilus ducreyi.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF219260; AAF33777.1; -.
KW   Transferase; Acyltransferase.
SQ   SEQUENCE   318 AA;  36799 MW;  E4C1F82FB88BE37A CRC64;


alignment_scores:
            Quality:  176.00              Length:    263
              Ratio:  1.128                 Gaps:     10
 Percent Similarity:  59.316   Percent Identity: 24.715

alignment_block:
 US-09-303-518D-571 x Q9L7U9    ..
```

```
Align seg 1/1  to: Q9L7U9  from: 1  to: 318

     132 TCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATA 181
         |||:::   |||      ||||||:::   :::   :::||| |||::::
      56 SerAsnTyrLeuLysLysLysGlyLysLysGlnTyrTyrArg.AlaAspI 72

     182 TGCGGCAGGCGGGTTTGAACCCCGAC.................ACGCAG 213
         :::::   :::          |||:::                 :::|||
      72 leAsnLeuCysTyrCysPheProGluTrpSerAlaAspLysArgAlaGln 88

     214 ACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGC 263
         :::::::::   :::|||   :::::::|||:::      |||   :::::
      89 ValIleGluGlnMetPheIleAsnValAlaGlnThrMetLeuGlyIleGl 105

     264 CCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCGG 313
         :          ::::::          ::::::          :::   :
     105 yGluIleAlaValArgSerValAlaHisLeuGlnLysArgSerGluPheI 122

     314 TACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTG 363
         ::   |||   ::::::::::::|||||||      ||||||:::::::::
     122 le...GlyLeuAsnTyrLeuLysGlnAlaLysAlaLysGlyLysAsnIle 137

     364 CTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACAT 413
         ::::::::::::|||||      |||   |||   :::|||      ||
     138 IleLeuLeuValProHisThrTrpSerIleAspAlaSerGly...IleIl 153

     414 CAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAA 463
         |   :::          :::   :::::::::|||||:::|||   :::
     153 eLeuHisThrTyrGlyMetProMetValSerMetTyrAsnProHisArgA 170

     464 TCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGC 513
                 :::|||   :::   :::|||   |||   |||      |||
     170 snAlaLeuValAspTrpLeuTrpAsnAlaThrArgGluArgPheGlyGly 186

     514 AAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCT 563
         |||         |||      :::|||:::|||   :::::::::   ::
     187 Lys...MetHisThrArgGlnAsnGlyIleLysProPheLeuAsnAspIl 202

     564 GCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGC 613
         :|||   |||:::         :::||||||||||                :
     202 eArgLysGlyAsnMetGlyTyrPheLeuProAsp.............G 214

     614 AGGAAGGCGGC.........GGCGTGTGGGCGGATTTTTTCGGCAAACCT 654
         :::::   |||         :::|||:::|||||||||||||:::
     214 luAspPheGlyGluGluLeuSerValTyrAlaAspPhePheAlaThrGlu 230

     655 GCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAA 704
               |||:::          |||:::|||:::|||   ::::::::
     231 LysAlaThrLeuProGlyLeuAsnLysMetAlaArgValAlaAsnAlaGl 247

     705 AACCCTG...TTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCG 751
         :::::::   :::|||         :::         |||      :::
     247 uValIleProMetPheProIleTyrAsnAlaGluLysGlyVal...TyrG 263

     752 TGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGCCCAC 801
           :::::::|||   |||:::      |||:::::::|||:::::|||:::
     263 lnMetGluIleLeuProAlaIle...GluPheSerGlySerSerAlaGln 278

     802 GATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTCC 851
         :::|||      :::|||:::         ||||||:::::::::::::   ||
     279 SerAlaArgGluMetAsnLysValIleGluTyrPheValHisLysAsnPr 295

     852 GACGCAGTATCTGTTTATGTACAACCGCTATAAAACG 888
         |:::||||||:::::::::   :::      ||||||
     295 oAlaGlnTyrValTrpIleLeuArgLeuLeuLysThr 307

seq_name: sp_bacteriap:O84013

seq_documentation_block:
ID   O84013      PRELIMINARY;      PRT;   455 AA.
AC   O84013;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
```

```
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ACYLTRANSFERASE.
GN   HTRB OR CT010.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
DR   EMBL; AE001275; AAC67600.1; -.
KW   Transferase; Acyltransferase; Complete proteome.
SQ   SEQUENCE   455 AA;  52058 MW;  0404B6326C67ACCF CRC64;


alignment_scores:
            Quality:  175.00              Length:    341
              Ratio:  1.054                Gaps:     12
 Percent Similarity:  48.680   Percent Identity: 24.633

alignment_block:
 US-09-303-518D-571 x O84013   ..

 Align seg 1/1  to: O84013  from: 1  to: 455

      22 CTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCCTG...  69
         ||||||      ||||||||||:::      |||||||:::      |||
       2 LeuPheLysArgLeuArgThrGlyGlyLysIleLeuValAspHisLeuVa  18

      70 .....................CTCAAATGCCTCTCCCTGCTGTCGCTTT  97
                              |||      |||   ||||||      |
      18 lTyrGlyLeuGlyLeuGlyValLeuThrIleLeuArgLeuLeuProArgS  35

      98 CCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTT 147
         ||   |||:::   :::::::::   ||||||         |||||||:::
      35 erSerLeuArgLeuPheSerLysGlyLeuGlyThrAlaLeuPheTyrPhe  51

     148 TTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATG.............. 183
         :::::::::   |||         ::::::|||:::
      52 IleSerAspValArgLysThrAlaLeuThrAsnLeuAlaLeuAlaPhePr  68

     184 ...................CGGCAGGCGGGTTTGAACCCCGACACGCAGA 214
                             |||            :::   ::::::|||
      68 oGluLysSerPheAlaGluArgTyrGlnIleAlaArgGlnSerValGlnG  85

     215 CGGTCAAAGCCGTTTTTGCGGAAACGGCAAAA.................. 246
           :::      :::|||:::|||   |||
      85 lnMetIleIleThrPheValGluLeuAlaThrValAspLysPheAlaLys 101

     247 .......................TGCGGTTTGGAACTTGCCCCC...GC 269
                                 :::      |||   ||||||   ::
     102 HisIleAspGluMetIleAlaIleAlaThrSerGluAspAlaProGluGl 118

     270 GTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCGGTACACG 319
         :||||||      ||||||:::::::::::::::::   :::
     118 yPhePhe......ProGluGluValSerSerGlnGlnGluLeu....... 130

     320 GCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTC 369
           :::|||      :::   |||||||:::   ||||||   ::::::
     131 .....AspHisPhePheSerArgLeuAspArgGlnGluGlyAlaIleLeu 145

     370 ATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCA 419
         :::      |||   :::::::::::::|||         ||||||:::::
     146 PheCysGlyHisGlnAlaAsnTrpGluLeuProPheLeuTyrIleThrLy 162

     420 GCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAG 469
         :::   |||           |||:::   ||||||   |||   :::
     162 sArgTyrProGly......LeuAlaPheAlaLysProValLysAsnArgA 177

     470 CGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACC 519
           :::::::::   :::   :::   |||         :::|||||||
     177 rgLeuAsnGlnLysIleIleSerLeuArgGluSerPheGlnGlyLysIle 193

     520 GCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGC 569
         :::|||         :::::::::::::|||   ::::::|||||||:::
     194 ValProProGln...AsnAlaIleAsnGlnAlaLeuArgAlaLeuHisAr 209

     570 GGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAG 619
          ||||||::::::   |||:::   |||:::|||         :::|||
     209 gGlyGluValValGlyIleValGlyAspGlnValLeuLeuSerSerGluT 226

     620 GCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACA 669
           :::                ::::|||||||:::   |||:::|||   |||
     226 yrSer.........TyrProLeuPheGlySerGlnAlaPheThrThrThr 239

     670 CTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTG 719
               |||   ||||||:::                  ::::::      ::
     240 SerProAlaLeuLeuAlaTyrLysThrLysLysThrValIleAlaValAl 256

     720 CTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCG 769
         :      |||   |||:::|||:::   :::|||:::
     256 aIleTyrArgLysProAsnGlyAsnTyrLeuValVal.............. 268

     770 TCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAAC 819
                                                 :::   :::|||:::
     269 ..............................ProSerLysAlaPheHis 274

     820 CGCAATACCGAATATTGGATACGC.......................... 843
           |||||||||       ||||||
     275 AlaAsnThrGluLeuSerIleArgGluSerThrGluGlnLeuMetAspAr 291

     844 .......CGTTTT.........................CCGACGCAGTATC 862
                 ||||||                      |||   |||:::|
     291 gLeuMetArgPheLeuGluLysGlyIleThrCysLysProGluGlnTrpL 308

     863 TGTTTATGTACAACCGCTATAAA 885
         ||:::::::::::::|||:::|||
     308 euTrpLeuHisLysArgTrpLys 315

seq_name: sp_bacteriap:Q9PL30

seq_documentation_block:
ID   Q9PL30      PRELIMINARY;      PRT;   453 AA.
AC   Q9PL30;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE, PUTATIVE.
GN   TC0278.
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MOPN / NIGG;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
DR   EMBL; AE002295; AAF39146.1; -.
DR   TIGR; TC0278; -.
KW   Complete proteome.
SQ   SEQUENCE   453 AA;  51463 MW;  6221515A00A093FF CRC64;
```

```
alignment_scores:
           Quality:  162.00              Length:    342
             Ratio:  0.970                Gaps:     13
 Percent Similarity:  48.830   Percent Identity: 23.977

alignment_block:
 US-09-303-518D-571 x Q9PL30   ..

 Align seg 1/1  to: Q9PL30  from: 1  to: 453

      22 CTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTG..... 60
         ||||||      ||||||::::::      :::|||:::
       2 LeuPheLysMetLeuArgSerGlyGlyLysValLeuValAspHisLeuVa 18

      61 .............ACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTT 97
                      ::::::|||       |||   ||||||      |
      18 lTyrGlyLeuGlyLeuGlyIleLeuThrIleLeuArgLeuLeuProArgS 35

      98 CCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTT 147
         ||   |||:::   :::|||:::   ||||||         |||||||::
      35 erSerLeuGlnLeuPheGlyLysGlyLeuGlyThrThrIlePheTyrVal 51

     148 TTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGG..... 186
         :::::::::   |||         :::::::|||:::
      52 IleSerAspPheArgLysThrAlaLeuThrAsnLeuAlaLeuAlaPhePr 68

     187 ........................CAGGCGGGTTTGAACCCCGACACGC 211
                                 |||   :::|||:::   ::::::|
      68 oGluLysSerPheThrGluArgTyrGlnIleAlaLeuLys...SerValG 84

     212 AGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAA..... 246
         ||   |||       :::|||   |||   |||
      84 lnGlnValIleIleThrPheIleGluLeuAlaThrValAspLysPheAla 100

     247 ..........................TGCGGTTTGGAACTTGCCCCC.. 267
                                  ::::::   |||   ||||||
     101 LysHisIleAspGluIleIleThrIleAlaSerSerGluAspAlaProGl 117

     268 .GCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCGGTAC 316
          :::||||||      ||||||:::::::::::::::   :::
     117 uGlyPhePhe......ProGluGluValSerSerGlnGlnGluLeu.... 130

     317 ACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTG 366
                ::::::      :::   |||||||:::   ||||||   :::
     131 ........AsnAsnPhePheSerArgLeuAspGlnGlnGluGlyAlaIle 144

     367 TTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAG 416
         ::::::      |||   :::::::::::|||         |||||::
     145 LeuPheCysGlyHisGlnAlaAsnTrpGluLeuProPheLeuTyrIleTh 161

     417 CCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCA 466
         :::::::   |||            |||:::   ||||||   |||
     161 rLysArgTyrProGly......LeuAlaPheAlaLysProValLysAsnP 176

     467 AAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAA 516
               :::::::::   :::   :::   |||         :::||||||
     176 roArgLeuAsnArgLysIleIleSerLeuArgGluSerPheGlnGlyLys 192

     517 ACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCG 566
            :::|||         ::::::::::::|||   ::::::|||||::
     193 IleValProProGln...AsnAlaIleAsnGlnAlaLeuArgAlaLeuHi 208

     567 CGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGG 616
         :   ||||||::::::   |||:::   |||:::|||         ::::
     208 sLysGlyGluValValGlyIleValGlyAspGlnValLeuLeuSerSerG 225

     617 AAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATG 666
         ::   :::                  :::|||||||:::   |||:::|||
     225 lnTyrSer.........TyrProLeuPheGlySerGlnAlaPheThrThr 238

     667 ACACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTT 716
         |||       |||   ||||||:::               ::::::
     239 ThrSerProAlaLeuLeuAlaTyrLysThrArgLysProValIleAlaVa 255

     717 CTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCC 766
          :::       |||   |||:::|||:::   :::|||:::
     255 lAlaIleTyrArgGlnProAsnGlyAsnTyrLeuValVal..... 268

     767 CCGTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTC 816
                                                 :::   :::|||
     269 ................................ProSerLysAlaPhe 273

     817 AACCGCAATACCGAATATTGGATACGC..... 843
               |||||||||       ||||||
     274 TyrAlaAsnThrGluLeuSerIleArgGluSerThrGluGlnLeuMetAs 290

     844 ..........CGTTTT.....................CCGACGCAGT 859
                     ||||||                  |||   |||:
     290 pLysLeuMetArgPheLeuGluLysGlyIleAlaCysLysProGluGlnT 307

     860 ATCTGTTTATGTACAACCGCTATAAA 885
         ::|||:::::::::::::|||:::|||
     307 rpLeuTrpLeuHisLysArgTrpLys 315

seq_name: sp_bacteriap:Q9ZCL1

seq_documentation_block:
ID   Q9ZCL1      PRELIMINARY;      PRT;   290 AA.
AC   Q9ZCL1;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE (HTRB).
GN   RP718.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
DR   EMBL; AJ235273; CAA15149.1; -.
KW   Complete proteome.
SQ   SEQUENCE   290 AA;  33809 MW;  991FF50AB841D5B3 CRC64;

alignment_scores:
           Quality:  159.00              Length:    205
             Ratio:  1.214                Gaps:      9
 Percent Similarity:  63.902   Percent Identity: 21.951

alignment_block:
 US-09-303-518D-571 x Q9ZCL1   ..

 Align seg 1/1  to: Q9ZCL1  from: 1  to: 290

     289 GACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCA 338
         ::::::|||:::   ::::::   :::   |||   |||::::::::::
      95 GluLeuGluSerArgIleGluIleIle...GlyLysGluAsnIleLysLy 110

     339 GGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCA 388
         :      |||       :::   :::|||:::::::::   |||:::::::
     110 sLeuGluAsp......GlnProPheLeuLeuPheSerGlyHisPheAlaA 125

     389 GCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTG 438
         ::::::|||::::::   :::   :::::::::       |||      :::
     125 snTrpAspIleSerLeuLysValLeuLysLysTyrTyrPro...LysVal 140
```

```
    439 ACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCA 488
        :::::::::|||:::      :::         :::::::|||:::::::
    141 AlaValIleTyrArgLysAlaAsnAsnProTyrValAsnLysLeuValAs 157

    489 GGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAG 538
        :   :::|||:::      |||   :::      |||   |||   :::|
    157 nGluSerArgAlaGlyAspLysLeuArgLeuIleProLysGlySerGluG 174

    539 GGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATC 588
        ||:::::::::::::::::|||::::::   ||||||:::   ::::::
    174 lyIleArgSerLeuValArgAlaIleLysGluGlyGluSerIleValMet 190

    589 CTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGA 638
        |||   |||:::            :::   :::   |||:::   :::
    191 LeuValAspGln.............LysMetAsnAspGlyIleGluValPr 203

    639 TTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCAC 688
         |||:::|||:::|||:::   |||   :::      ||||||:::|||
    203 oPheLeuGlyArgProSerMetThrAlaSerSerIleAlaLysIleAla. 219

    689 ACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTG...CCC 735
                    |||      ::::::   |||:::   :::
    220 .....LeuGlnTyrLysTyrProIleIleProCysGlnIleIleArgThr 234

    736 GACGGACAAGGCTTCGTGTTGCACATCCGCCCC...GTCCAAGGGGAATT 782
           |||:::   |||   :::   ::::::|||   ::::::   |||
    235 LysGlySerTyrPheLysValIleValHisProGlnLeuLysPheGluGl 251

    783 GAACGGCAACAAAGCCCACGATGCC......GCCGTGTTCAACCGCAAT. 825
         :::|||::::::      |||:::         :::|||||   |||
    251 nThrGlyAspAsnLysValAspCysTyrAsnIleMetPheAsnIleAsnG 268.

    826 .....ACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATG 870
                    |||:::::::::   |||:::|||::::::::::::
    268 lnIleLeuGlyGluTrpValLysGlnAsnProAlaGlnTrpPheTrpPhe 284

    871 TACAACCGCTATAAA 885
        :::|||||||:::|||
    285 HisAsnArgTrpLys 289

seq_name: sp_bacteria:O82937

seq_documentation_block:
ID   O82937      PRELIMINARY;      PRT;   332 AA.
AC   O82937;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LIPID A BIOSYNTHESIS (KDO)2-(LAUROYL)-LIPID IVA ACYLTRANSFERASE.
GN   ECF4 OR MSBB.
OS   Escherichia coli O157:H7.
OG   Plasmid pO157.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98290540; PubMed=9628576;
RA   Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,
RA   Yatsudo H.C., Kubota Y., Yamaichi Y., Iida T., Yamamoto K., Honda T.,
RA   Han C., Ohtsubo A., Kasamatsu M., Hayashi T., Kuhara S., Shinagawa H.;
RT   "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
RT   enterohemorrhagic Escherichia coli O157:H7 derived from Sakai
RT   outbreak.";
RL   DNA Res. 5:1-9(1998).
RN   [2]
RP   SEQUENCE OF 1-235 FROM N.A.
RC   STRAIN=4304-PT14;
RX   MEDLINE=98261495; PubMed=9596716;
RA   Boerlin P., Chen S., Colbourne J.K., Johnson R., De Grandis S.,
RA   Gyles C.;
RT   "Evolution of enterohemorrhagic Escherichia coli hemolysin plasmids
RT   and the locus for enterocyte effacement in shiga toxin-producing E.
RT   coli.";
RL   Infect. Immun. 66:2553-2561(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EDL933;
RA   Brunder W.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011549; BAA31840.1; -.
DR   EMBL; AF043470; AAC24348.1; -.
DR   EMBL; Y11275; CAA72141.1; -.
KW   Transferase; Acyltransferase; Plasmid.
SQ   SEQUENCE   332 AA;  37858 MW;  0D49F6B93E29E9A6 CRC64;


alignment_scores:
            Quality:  156.50              Length:    273
              Ratio:  1.057                 Gaps:     10
 Percent Similarity:  54.212   Percent Identity: 23.810

alignment_block:
 US-09-303-518D-571 x O82937   ..

 Align seg 1/1  to: O82937  from: 1  to: 332

    121 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCAT 170
        :::||||||   |||:::      |||   |||:::   |||   |||
     69 LysLeuGlyMetLeuValGlyArgLeuGlyLysSerAlaArgGlnArgAl 85

    171 CGTCGCCAATATGCGG.....CAGGCGGGTTTGAACCCCGACA 208
         :::   |||:::            :::   :::   :::
     85 aLeuIleAsnLeuSerLeuCysPheProGluTyrSerAspLysGluLysG 102

    209 CGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAA 258
         :::   |||   |||:::||||||   ::::::   :::   :::
    102 luAsnIleValAspAlaMetPheAlaThrAlaSerMetAlaValValLeu 118

    259 CTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAA 308
        :::|||         ::::::   |||:::   |||:::      :::::
    119 MetAlaGluLeuAlaLeuSerGlyProAspLysIleSerHisArgIleAr 135

    309 AGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAG 358
        :      :::|||   |||   |||:::::      ::::::::::|||
    135 g...TrpAsnGlyLeuGluIleValGluLysMetAlaGlnAsnAsnGluL 151

    359 GGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTG..... 399
        ::::::|||::::::|||||      :::   |||:::
    151 ysValIlePheLeuValProHisAlaTrpGlyValAspIleProAlaMet 167

    400 ............GGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCT 437
                      :::|||||||                      ::
    168 LeuMetAlaAlaSerGlyArg.....................LysMe 176

    438 GACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGC 487
        ::::|||||||:::         :::      :::::|||   :::   :
    176 tAlaAlaMetPheHisAsnGlnArgAsnProValValAspTyrValTrpA 193

    488 AGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAA 537
        :::::   |||   |||      ||||||            :::   :::
    193 snSerValArgArgArgPheGlyGlyLysLeuHisAlaArgAsn...Asp 208

    538 GGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCAT 587
        |||:::   :::::::::::::::::|||   |||
    209 GlyIleAlaSerPheValArgSerValArgGlnGlyTyrTrpGlyTyrTy 225

    588 CCTGCCC......GACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGT 631
         ||||||      ||||||   |||:::   :::|||          :
    225 rLeuProAspGlnAspHisGlyProGluPheSerGlu.....P 238

    632 GGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAA 681
        ::|||||||||||||:::            |||:::   :::   ::::::
    238 heAlaAspPhePheAlaThrTyrLysAlaThrLeuProValIleGlyArg 254

    682 TTGGCACACGTCAAAGGCGTGAAA......ACCCTGTTTTTCTGCTGCGA 725
```

```
      |||:::::::::::::|||::::::         ||||||
  255 LeuSerArgIleSerGlyAlaArgIleIleProLeuPhe..... 267

  726 ACGCCTGCCCGACGGACAA......GGCTTCGTGTTGCACATCCGCCCCG 769
        :::   ||||||:::         :::::::::|||:::   |||
  268 .ProValTyrAspGlyLysThrHisHisLeuThrIleHisValSerProP 284

  770 TCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAAC 819
        :::      :::   :::   ||||||   |||      :::|||
  284 roLeuAlaIleArgGlnLysSerAspAlaHisIleAlaArgGlnIleAsn 300

  820 CGCAATACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTAT 869
      :::   :::|||   ::::::|||      |||   ||||||   :::::
  301 GluValValGluAsnPheValArgProHisProGluGlnTyrThrTrpIl 317

  870 GTACAACCGCTATAAAACG 888
      :   :::      ||||||
  317 eLeuLysLeuLeuLysThr 323

seq_name: sp_bacteria:Q9ZGU3

seq_documentation_block:
ID   Q9ZGU3      PRELIMINARY;      PRT;   343 AA.
AC   Q9ZGU3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE ACYLTRANSFERASE.
GN   L7029.
OS   Escherichia coli O157:H7.
OG   Plasmid pO157.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EDL933;
RX   MEDLINE=98391744; PubMed=9722640;
RA   Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA   Blattner F.R.;
RT   "The complete DNA sequence and analysis of the large virulence plasmid
RT   of Escherichia coli O157:H7.";
RL   Nucleic Acids Res. 26:4196-4204(1998).
DR   EMBL; AF074613; AAC70097.1; -.
KW   Transferase; Acyltransferase; Plasmid.
SQ   SEQUENCE   343 AA;  39074 MW;  EB869F0D3D30DB6A CRC64;


alignment_scores:
            Quality:  156.50              Length:    273
              Ratio:  1.057                 Gaps:     10
 Percent Similarity:  54.212   Percent Identity: 23.810

alignment_block:
 US-09-303-518D-571 x Q9ZGU3   ..

 Align seg 1/1  to: Q9ZGU3  from: 1  to: 343

  121 CGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCAT 170
      :::||||||   |||:::      |||   |||:::   |||   |||
   80 LysLeuGlyMetLeuValGlyArgLeuGlyLysSerAlaArgGlnArgAl 96

  171 CGTCGCCAATATGCGG............CAGGCGGGTTTGAACCCCGACA 208
        :::   |||:::              :::   :::   :::
   96 aLeuIleAsnLeuSerLeuCysPheProGluTyrSerAspLysGluLysG 113

  209 CGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAA 258
        :::   |||   |||:::||||||   ::::::   :::   :::
  113 luAsnIleValAspAlaMetPheAlaThrAlaSerMetAlaValValLeu 129

  259 CTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAA 308
      :::|||        ::::::   |||:::   |||:::      :::::
  130 MetAlaGluLeuAlaLeuSerGlyProAspLysIleSerHisArgIleAr 146

  309 AGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAG 358
      :      :::|||   |||   |||:::::      :::::::::|||
  146 g...TrpAsnGlyLeuGluIleValGluLysMetAlaGlnAsnAsnGluL 162

  359 GGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTG......... 399
        ::::::|||:::::::||||||      :::   |||:::
  162 ysValIlePheLeuValProHisAlaTrpGlyValAspIleProAlaMet 178

  400 ............GGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCT 437
                  :::|||||||                          ::
  179 LeuMetAlaAlaSerGlyArg.......................LysMe 187

  438 GACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGC 487
      ::::|||||||:::         :::      :::::|||   :::   :
  187 tAlaAlaMetPheHisAsnGlnArgAsnProValValAspTyrValTrpA 204

  488 AGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAA 537
      :::::   |||   |||      ||||||              :::   :::
  204 snSerValArgArgArgPheGlyGlyLysLeuHisAlaArgAsn...Asp 219

  538 GGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCAT 587
      |||:::   ::::::::::::::::::|||   |||
  220 GlyIleAlaSerPheValArgSerValArgGlnGlyTyrTrpGlyTyrTy 236

  588 CCTGCCC......GACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGT 631
       ||||||      ||||||   |||:::   :::|||           :
  236 rLeuProAspGlnAspHisGlyProGluPheSerGlu.....P 249

  632 GGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAA 681
      ::|||||||||||||:::            |||:::   :::   ::::::
  249 heAlaAspPhePheAlaThrTyrLysAlaThrLeuProValIleGlyArg 265

  682 TTGGCACACGTCAAAGGCGTGAAA......ACCCTGTTTTTCTGCTGCGA 725
      |||:::::::::::::|||::::::      ||||||
  266 LeuSerArgIleSerGlyAlaArgIleIleProLeuPhe..... 278

  726 ACGCCTGCCCGACGGACAA......GGCTTCGTGTTGCACATCCGCCCCG 769
        :::   ||||||:::         :::::::::|||:::   |||
  279 .ProValTyrAspGlyLysThrHisHisLeuThrIleHisValSerProP 295

  770 TCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAAC 819
        :::      :::   :::   ||||||   |||      :::|||
  295 roLeuAlaIleArgGlnLysSerAspAlaHisIleAlaArgGlnIleAsn 311

  820 CGCAATACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTAT 869
      :::   :::|||   ::::::|||      |||   ||||||   :::::
  312 GluValValGluAsnPheValArgProHisProGluGlnTyrThrTrpIl 328

  870 GTACAACCGCTATAAAACG 888
      :   :::      ||||||
  328 eLeuLysLeuLeuLysThr 334

seq_name: sp_bacteria:Q9AKA6

seq_documentation_block:
ID   Q9AKA6      PRELIMINARY;      PRT;   290 AA.
AC   Q9AKA6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HTRB PROTEIN.
GN   HTRB.
OS   Rickettsia typhi.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=785;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WILMINGTON;
RX   MEDLINE=21219194; PubMed=11319266;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Pseudogenes, junk DNA and the dynamics of Rickettsia genomes.";
```

```
RL   Mol. Biol. Evol. 18:829-839(2001).
DR   EMBL; AJ293328; CAC33762.1; -.
SQ   SEQUENCE   290 AA;  33814 MW;  2E9B2D3193B70EDD CRC64;


alignment_scores:
            Quality:  155.50              Length:    195
              Ratio:  1.275                 Gaps:      7
 Percent Similarity:  62.564   Percent Identity: 22.051

alignment_block:
 US-09-303-518D-571 x Q9AKA6   ..

 Align seg 1/1  to: Q9AKA6  from: 1  to: 290

     319 GGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTT 368
         |||   |||::::::|||:::      |||       :::   :::|||::
     104 GlyLysGluAsnIleGlnLysLeuGluAsp......GlnProPheLeuLe 118

     369 CATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCC 418
         :::::::   |||:::::::::::::|||::::::   :::   :::::::
     118 uPheSerGlyHisPheAlaAsnTrpAspIleSerLeuLysValLeuLysL 135

     419 AGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAA 468
         ::      |||      :::::::::::::|||:::      :::
     135 ysTyrTyrPro...LysValAlaValIleTyrArgLysAlaAsnAsnPro 150

     469 GCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAAC 518
         ::::::|||:::::::::   :::|||:::      |||   :::
     151 TyrValAsnLysLeuValAsnGluSerArgAlaGlyAspLysLeuArgLe 167

     519 CGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCG 568
           |||   |||   :::|||::::::   :::::::::|||::::::
     167 uIleProLysGlySerGluGlyIleArgAlaLeuValSerAlaIleLysG 184

     569 CGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAA 618
         ||||||:::   ::::::|||   |||:::          :::
     184 luGlyGluSerIleValMetLeuValAspGln...........LysMet 196

     619 GGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGAC 668
         :::   |||:::   :::   |||:::|||:::|||:::   |||   ::
     197 AsnAspGlyIleGluValProPheLeuGlyArgProSerMetThrAlaSe 213

     669 ACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCT 718
         :      ||||||:::|||          |||       ::::::
     213 rAlaIleAlaLysIleAla......LeuGlnTyrLysTyrProIleIleP 228

     719 GCTGCGAACGCCTG...CCCGACGGACAAGGCTTCGTGTTGCACATCCGC 765
           |||:::   :::         |||:::   |||   :::   ::::::
     228 roCysGlnIleIleArgThrLysGlySerTyrPheLysValIleValHis 244

     766 CCC...GTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCC..... 807
         |||   ::::::   |||   :::|||::::::      |||:::
     245 ProGlnLeuLysPheGluGlnThrGlyAspAsnLysValAspCysTyrAs 261

     808 .......GCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTC 850
                    :::|||:::          |||::::::::::   |
     261 nIleMetPheTyrIleAsnGlnIleLeuGlyGluTrpValLysGlnAsnP 278

     851 CGACGCAGTATCTGTTTATGTACAACCGCTATAAA 885
         ||:::|||:::::::::::::::|||||:::|||
     278 roAlaGlnTrpPheTrpPheHisAsnArgTrpLys 289

 seq_name: sp_bacteriap:Q9CME5

 seq_documentation_block:
 ID   Q9CME5      PRELIMINARY;      PRT;   313 AA.
 AC   Q9CME5;
 DT   01-JUN-2001 (TrEMBLrel. 17, Created)
 DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE   HYPOTHETICAL PROTEIN PM0885.
 GN   PM0885.

OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR   EMBL; AE006127; AAK02969.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   313 AA;  35956 MW;  23456DB87C45E535 CRC64;


alignment_scores:
            Quality:  154.00              Length:    287
              Ratio:  0.975                 Gaps:     10
 Percent Similarity:  55.052   Percent Identity: 25.087

alignment_block:
 US-09-303-518D-571 x Q9CME5   ..

 Align seg 1/1  to: Q9CME5  from: 1  to: 313

      64 GCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCT 113
         ||||||   :::   :::          |||            ||||||   ||
      26 AlaLeuTrpArgLeuIleValCysLeuProTyrProIleLeuHisLysLe 42

     114 GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCG 163
         |||||||   ||||||   |||         |||   |||:::      :
      42 uGlyAsnGlyLeuGlyTrpLeu.........LeuAlaLysSerThrValG 56

     164 CGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAAC..... 201
         :::::      ||||||:::   :::   :::|||:::
      56 lyGlnArgArgAlaAsnIleAlaArgArgAsnLeuGlnLeuCysPhePro 72

     202 .....CCCGACACGCAGACG...GTCAAAGCCGTTTTTGCGGA 236
                  :::   :::|||  :::   |||   :::
      73 HisTyrSerAlaAspGlnIleGluThrLeuLeuHisAlaAsnLeuLysAl 89

     237 AACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGG 286
         |||:::   :::   :::|||   :::   |||:::|||   :::
      89 aThrGlyMetAlaIleIleGluThrGlyMetAlaTrpPheTrpSerAspA 106

     287 AAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAG 336
           |||:::          |||   ::::::|||   |||::::::::::
     106 laArgIleLysLysTrpSerLys...IleGluGlyLeuGluTyrLeuLys 121

     337 CAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGG 386
         :::         |||   :::|||::::::::::::      |||:::
     122 GluTyr......LysGlnAspGlyValIleLeuValGlyValHisPheLe 136

     387 CAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACC 436
          :::   :::||||||:::|||   ::::::   :::   |||
     136 uThrLeuGluLeuGlyAlaArgIleValGlyLeuHisGlnProGly.... 151

     437 TGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATG 486
           ::::::|||:::|||                ::::|||   ::::::
     152 ..IleGlyValTyrArgProAsnAspAsnProLeuMetAspTrpLeuGln 167

     487 CAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACA 536
           ||||||:::|||::::::                          ::
     168 ValGlnGlyArgLeuArgSerAsnLysAspLeuLeuAspArg......Ly 182

     537 AGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCA 586
         :   ::::::   :::|||:::|||||||||   ||||||
     182 sAspLeuArgGlyMetIleArgAlaLeuArgGlnGlyGluIleIleTrpT 199

     587 TCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCG 636
           |||||||||      :::      :::   ::::::|||::::::
     199 yrAlaProAspHis......AspTyrGlyArgLysAsnAlaValPheVal 213
```

```
     637 GATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCA......AA 680
         ||||||:::   |||      :::   |||   ::::::
     214 ProPhePheAlaValProAspAlaAlaThrThrThrGlySerTyrTyrLe 230

     681 ATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCC 730
         |||                 |||::::::   |||:::      |||
     230 uLeuLysSerSerProThrSerLysValIleProPheAlaProLeuArgA 247

     731 TGCCCGACGGACAAGGCTTCGTGTTGCACATCCGC...CCCGTCCAAGGG 777
             ||||||:::|||:::::::::   |||      ||||||:::
     247 snSerAspGlySerGlyTyrThrValLysIleSerProProValAspPhe 263

     778 GAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAATAC 827
         :::         ::::::::::::   |||:::   :::|||:::   ::
     264 SerAspIleThrAspGluThrArgIleAlaThrArgMetAsnGlnValVa 280

     828 CGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACC 877
         :|||      |||   :::         ||||||:::::::::::::::|
     280 lGluThrGluIleMetLysGlyValGluGlnTyrMetTrpLeuHisArgA 297

     878 GCTATAAAACG 888
         ||:::||||||
     297 rgPheLysThr 300

seq_name: sp_bacteria:Q9AKF0

seq_documentation_block:
ID   Q9AKF0      PRELIMINARY;      PRT;   290 AA.
AC   Q9AKF0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HTRB PROTEIN.
GN   HTRB.
OS   Rickettsia rickettsii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=783;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=84-21C;
RX   MEDLINE=21219194; PubMed=11319266;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Pseudogenes, junk DNA and the dynamics of Rickettsia genomes.";
RL   Mol. Biol. Evol. 18:829-839(2001).
DR   EMBL; AJ293329; CAC33715.1; -.
SQ   SEQUENCE   290 AA;  33550 MW;  9FCC5D73E5CBE89A CRC64;


alignment_scores:
           Quality:  149.50            Length:    198
             Ratio:  1.225               Gaps:      7
 Percent Similarity:  61.616   Percent Identity: 22.222

alignment_block:
 US-09-303-518D-571 x Q9AKF0   ..

 Align seg 1/1  to: Q9AKF0  from: 1  to: 290

     310 GCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGG 359
         |||:::   |||   |||:::::::::::::         :::|||:::
     101 AlaIleIleGlyIleGluAsnIleLysLysLeu......GluGlyGlnPr 115

     360 GCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCT 409
         :::|||:::::::::   |||:::::::::::|||::::::   :::
     115 oPheLeuLeuPheSerGlyHisPheAlaAsnTrpAspIleSerIleLysI 132

     410 ACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCG 459
         :::   :::      |||      ::::::::::::|||:::
     132 leLeuHisLysPheTyrPro...LysValAlaValIleTyrArgLysAla 147

     460 AAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAA 509
         :::         ::::::|||:::::::::   :::|||:::      ||
     148 AsnAsnProTyrValAsnLysLeuValAsnGluSerArgAlaGlyAspLy 164

     510 AGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGG 559
         |   :::      |||   |||   :::|||   :::   :::::::::|
     164 sLeuArgLeuIleProLysGlyProGluGlySerArgAlaLeuValArgA 181

     560 CCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCT 609
         ||::::::   :::||||||   ::::::|||   |||:::
     181 laIleLysGluSerGluAlaIleValMetLeuValAspGln......... 194

     610 CCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATA 659
            :::   :::   |||:::   :::   |||:::|||   ||||||
     195 ...LysMetAsnAspGlyIleGluValProPheLeuGlyHisProAlaMe 210

     660 CACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCC 709
         |||   :::      ||||||:::|||              |||
     210 tThrAlaSerAlaIleAlaLysIleAla......LeuGlnTyrLysTyrP 225

     710 TGTTTTTCTGCTGCGAACGCCTG...CCCGACGGACAAGGCTTCGTGTTG 756
           ::::::   |||:::   :::          |||:::   |||   :::
     225 roIleIleProCysGlnIleIleArgThrLysGlySerTyrPheLysVal 241

     757 CACATCCGCCCC...GTCCAAGGGGAATTGAACGGCAACAAAGCCCACGA 803
            ::::::|||   ::::::   |||   :::|||::::::      ||
     242 IleValHisProGlnLeuLysLeuGluGlnThrGlyAspAsnLysAlaAs 258

     804 TGCC............GCCGTGTTCAACCGCAATACCGAATATTGGATAC 841
         |:::                ::::|||:::            |||::::
     258 pCysTyrAsnIleMetLeuAsnIleAsnGlnMetLeuGlyGluTrpValL 275

     842 GCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAA 885
         :::::   |||:::|||:::::::::::::::::|||||::::|||
     275 ysGlnAsnProAlaGlnTrpPheTrpPheHisAsnArgTrpLys 289

seq_name: sp_bacteria:P94909

seq_documentation_block:
ID   P94909      PRELIMINARY;      PRT;   529 AA.
AC   P94909;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 58.7 KDA PROTEIN (FRAGMENT).
OS   Microbacterium ammoniaphilum.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Micrococcineae; Microbacteriaceae; Microbacterium.
OX   NCBI_TaxID=42460;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15354;
RX   MEDLINE=96257250; PubMed=8654988;
RA   Striebel H.M., Seeber S., Jarsch M., Kessler C.;
RT   "Cloning and characterization of the MamI restriction-modification
RT   system from Microbacterium ammoniaphilum in Escherichia coli.";
RL   Gene 172:41-46(1996).
DR   EMBL; X79027; CAA55649.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   529 AA;  58716 MW;  5DC1216237DF772D CRC64;


alignment_scores:
           Quality:  146.00            Length:    320
             Ratio:  1.090               Gaps:     19
 Percent Similarity:  41.875   Percent Identity: 27.188

alignment_block:
 US-09-303-518D-571 x P94909   ..

 Align seg 1/1  to: P94909  from: 1  to: 529

      30 CCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCCTGCTCAAATGCC 79
```

```
   |||      :::|||::::::|||||::::::|||||   |||   ||
 74 ProAlaGlnArgArgGlnValHisProLeuGlnArgProHisGlnGluPr 90

 80 T..............CTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTG 114
    |              ||||||  |||           |||:::   |
 90 oGlyAlaArgGlyGluLeuPro...ValArgAspAspArgAlaGluArgG 106

115 GGAAA......CCGGCTCGGACATCT..................... 134
    ||:::      |||||||||   :::
106 lyArgArgGluProAlaArgProAlaAlaArgAspAlaArgArgHisLeu 122

135 ......GGCGTTTTACCTTTT...AAAGGAAGACCGCGCGCGCATCGTCG 175
          |||:::   |||      :::|||      |||   ||||||||
123 ArgLeuGlyAlaHisProAlaArgArgGlyValLeuArgArgHisArgAr 139

176 CCAATATGCGGCAGGC......GGGTTTGAACCCCGA............. 206
    |:::   |||   |||      |||      ||||||
139 gHisArgAlaArgGlyGluArgGlyArgGlyProArgGlnGlnValProA 156

207 .....CACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGG 251
         |||      |||:::         :::|||
156 rgGlnHisProArgGlyArgArgAspArgAlaGly.............. 167

252 TTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAA 301
                              :::   |||   ||||||:::::
168 ........................ArgProGlyLeuHisArgArgAr 175

302 TGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAG 351
    :   ::::::      |||   |||      :::      |||:::
175 gArgArgAlaArgArgArgGlnGlyArgProGlnValArgHisGlyAspA 192

352 GGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGG 401
      :::            |||:::|||      |||:::   |||
192 spGln............HisArgAlaAspProArgArgProArgAspPro 204

402 CGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACA 451
    |||:::   |||:::|||                              ::
205 ArgAlaGlyHisHisProLeuArg.......................Gl 213

452 AGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCA...GGCGGGCAGG 498
    :::::::::::::|||:::   |||      :::|||   ||||||||||
213 uGlyGlyGlnGlyGlnGluAlaArgProLeuArgAlaArgGlyGlyGlnG 230

499 GT.............................................. 500
    ||
230 lyGlyProGlyGlyProAlaGlyArgHisProAlaLeuGlyAspGluAsp 246

501 .........GCGCGGCAAAGGCAAAACCGCGCCCACCGG.........CA 532
             |||||||||      :::   ||||||||         ||
247 ArgProArgAlaArgGlnGlyAlaArgAlaAlaHisArgGluAlaLeuHi 263

533 TACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACC 582
    |   |||   |||   :::                |||||||||::::::
263 sLeuArgLeuGlnArgArg..............ArgGlyArgAlaAspG 275

583 ATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTG 632
           |||:::   |||         |||   |||   ||||||:::
275 lyCysArgAlaGlnGlyArg.........AlaArgArgAlaArgArgAla 288

633 GGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAAT 682
    ||||||                                 ||||||
289 GlyGly................................GlyGly..... 292

683 TGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTG 732
                                                  |||
293 ..........................................LeuProA 295

733 CCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGA... 779
      |||:::   |||:::|||   ||||||   ||||||||||||||||
295 rgArgGluAspArgValArg...AlaHisArgProArgProArgGlyArg 310

780 ..............ATTGAACGGCAACAAAGCCCACGATGCCGCCGTGT 814
                  :::::::|||      :::      :::|||   ::
```

```
311 GlyGlyAlaAlaArgLeuAspArgAlaGlyGlyValGlyAlaArgProAl 327

815 TCAACCGCAA 824
    ::::|||:::
327 aArgProArg 330
```

seq_name: sp_invertebrate:Q9VU43

```
seq_documentation_block:
ID   Q9VU43      PRELIMINARY;      PRT;   954 AA.
AC   Q9VU43;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG11274 PROTEIN.
GN   CG11274.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003539; AAF49848.1; -.
DR   FlyBase; FBgn0036340; CG11274.
DR   InterPro; IPR002483; PWI.
DR   InterPro; IPR000230; Ribosomal_S12.
DR   Pfam; PF01480; PWI; 1.
DR   SMART; SM00311; PWI; 1.
DR   PROSITE; PS00055; RIBOSOMAL_S12; UNKNOWN_1.
SQ   SEQUENCE   954 AA;  107592 MW;  572216979110A107 CRC64;
```

```
alignment_scores:
            Quality:  145.00              Length:    329
              Ratio:  1.074                Gaps:     17
 Percent Similarity:  41.033   Percent Identity: 25.228

alignment_block:
 US-09-303-518D-571 x Q9VU43   ..

 Align seg 1/1  to: Q9VU43  from: 1  to: 954

     150 AAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGG..... 194
         :::|||:::   |||:::      ||||||::::::::||||||
     347 ArgGlyAsnAspArgGly......ArgGlnPheGlyGlyGlyGlyGl 361

     195 ..................TTTGAACCCCGACACGCAGAC......GGTC 219
                           |||:::||||||      :::       |
     361 yArgArgGlyGlyArgArgPheSerProArgGlyArgSerProMetArgG 378

     220 AAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTT............... 254
         ||:::::::      |||||||||      ||||||
     378 lnAspAsnGlyGlyGlyAsnGlyGlyAsnArgPheGlnArgSerAsnSer 394

     255 ..........................GGAACTTGCCCCCGCGTTTTTCA 277
                                    :::      |||      :::::
     395 ArgArgArgSerArgSerArgArgLeuSerArgSerProMetArgTyrSe 411

     278 AAAAACCGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAA 327
         :::::::   |||:::   ||||||   :::   :::       ::::::
     411 rArgSerProArgArgPheAsnAsnArgArgArgSerProMetMetAsnP 428

     328 CACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCC 377
           |||::::::|||   |||   ||||||                  :::
     428 heArgGlyGlyGlyArgGlyGlyGlyArg..............GlyGly 439

     378 GCACATCGGCAGCTA..............................CGATT 397
         :::||||||||||::                              |||
     440 GlyHisArgGlnMetTrpGlnHisArgGlyGlySerProAsnPheArgGl 456

     398 TGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCT.......... 437
          |||      :::|||      |||          |||
     456 yGlyGlyArgIleHisTrpGlnAlaArgHisSerProGlyGlyGlyGlyA 473

     438 ............................GACCGCCATGTACAAGCCGCC 458
                                     ||||||
     473 rgGlyGlyGlyGlyGlyMetGlyArgPheAspArg............... 484

     459 GAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCA 508
           :::||||||   :::||||||
     485 ...ArgGlnSerProGlnGlnAsn........................... 491

     509 AAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAAT....... 551
          |||   ||||||   ||||||::::::      |||   |||
     492 .ArgTyrAsnArgAspHisArgGlnSerProMetGlnGlnAsnGlnProP 508

     552 ...................CATCAAGGCCCTGCGCGCGGGCGAGGCAAC 581
                            ||||||||||||:::|||      ::::::
     508 heArgLysGlnPheSerProHisGlnGlyProGlyArgArgPheSerSer 524

     582 CATCATCCTGCCCGA............CCACGTCCCTTCTCCGCAGGAAG 619
           :::      |||              ||||||   |||      ||
     525 ProGlnGlnArgArgAsnSerArgAspArgArgProAsnSerArgGluAr 541

     620 GCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACA 669
         ||||         ||||||
     541 gArgSerSerProGlyGly............................... 547

     670 CTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTG 719
           ||||||      |||      :::|||   :::||||||
     548 ..GlyGly......GlyHisMetAsnArgTrpAspAsnPro......... 558

     720 CTGCGAACGCCTGCC................................... 734
                 ||||||
     559 ......ProProAlaArgAsnArgArgSerSerSerGlySerSerAlaGl 573

     735 ....CGACGGACAAGGCTTCGTGTTGCACATCCGCCC............. 767
             |||:::   |||   |||   |||   ||||||
     573 yGlyArgGlnGlnArgGlnArgSerAlaSerProProGluLysArgArgS 590

     768 ...........CGTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGC 806
                    |||   |||:::::::::|||::::::|||   |||
     590 erSerHisSerArgSerArgSerLeuSerArgAsnArgSer...ArgSer 605

     807 CGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTCCGACGC 856
         |||      :::   :::   |||      ::::::|||::::::   ::
     606 ArgSerSerSerGlnHisSerArgLysArgGluSerProIleGlyArgSe 622

     857 AGTATCTGTTTATGTACAACCGCTATAAAACGCCGTA 893
         :        |||:::   |||      |||::::::
     622 rSerValGluTyrAlaGlyProAlaValAsnThrIle 634

seq_name: sp_bacteriap:Q92GN2

seq_documentation_block:
ID   Q92GN2      PRELIMINARY;      PRT;   290 AA.
AC   Q92GN2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE.
GN   HTRB OR RC1091.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MALISH 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
DR   EMBL; AE008660; AAL03629.1; -.
KW   Transferase; Acyltransferase; Complete proteome.
SQ   SEQUENCE   290 AA;  33522 MW;  1C0FD05318332FD0 CRC64;


alignment_scores:
            Quality:  144.50              Length:    198
              Ratio:  1.194                Gaps:      7
 Percent Similarity:  61.111   Percent Identity: 21.717

alignment_block:
 US-09-303-518D-571 x Q92GN2   ..

 Align seg 1/1  to: Q92GN2  from: 1  to: 290

     310 GCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGG 359
         |||:::   |||   |||:::::::::::::          :::|||:::
     101 AlaIleIleGlyIleGluAsnIleLysLysLeu......GluGlyGlnPr 115

     360 GCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCT 409
          :::|||::::::      |||:::::::::::::|||::::::   :::
     115 oPheLeuLeuPheGlyGlyHisPheAlaAsnTrpAspIleSerIleLysI 132

     410 ACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCG 459
           :::   :::      |||      ::::::::::::|||:::
     132 leLeuHisLysPheTyrPro...LysValAlaValIleTyrArgLysAla 147

     460 AAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAA 509
         :::         ::::::|||:::::::::   :::|||:::      ||
     148 AsnAsnProTyrValAsnLysLeuValAsnGluSerArgAlaGlyAspLy 164

     510 AGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGG 559
         |   :::      |||   |||   :::|||   :::   :::::::::|
```

```
     164 sLeuArgLeuIleProLysGlyProGluGlySerArgAlaLeuValArgA 181

     560 CCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCT 609
         ||:::::   :::|||:::   :::::|||   |||:::
     181 laIleLysGluSerGluSerIleValMetLeuValAspGln......... 194

     610 CCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATA 659
            :::   :::   |||:::   :::   |||:::|||   ||||||
     195 ...LysMetAsnAspGlyIleGluValProPheLeuGlyHisProAlaMe 210

     660 CACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCC 709
          |||   :::      ||||||:::|||               |||
     210 tThrAlaSerAlaIleAlaLysIleAla......LeuGlnTyrLysTyrP 225

     710 TGTTTTTCTGCTGCGAACGCCTG...CCCGACGGACAAGGCTTCGTGTTG 756
         ::::::   |||:::   :::          |||:::   |||   :::
     225 roIleIleProCysGlnIleIleArgThrLysGlySerTyrPheLysVal 241

     757 CACATCCGCCCC...GTCCAAGGGGAATTGAACGGCAACAAAGCCCACGA 803
         :::::::|||   ::::::   |||   :::|||::::::      ||
     242 IleValHisProGlnLeuLysPheGluGlnThrGlyAspAsnLysAlaAs 258

     804 TGCC............GCCGTGTTCAACCGCAATACCGAATATTGGATAC 841
         |:::                  :::|||:::            |||::::
     258 pCysTyrAsnIleMetLeuAsnIleAsnGlnMetLeuGlyGluTrpValL 275

     842 GCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTATAAA 885
         :::::   |||:::|||:::::::::::::::::||||||:::|||
     275 ysGlnAsnProAlaGlnTrpPheTrpPheHisAsnArgTrpLys 289

seq_name: sp_bacteria:Q9AKK5

seq_documentation_block:
ID   Q9AKK5      PRELIMINARY;      PRT;   290 AA.
AC   Q9AKK5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HTRB PROETIN.
GN   HTRB.
OS   Rickettsia montana.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=33991;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21219194; PubMed=11319266;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Pseudogenes, junk DNA and the dynamics of Rickettsia genomes.";
RL   Mol. Biol. Evol. 18:829-839(2001).
DR   EMBL; AJ293330; CAC33650.1; -.
SQ   SEQUENCE   290 AA;  33635 MW;  9599E3C0C3C076F9 CRC64;


alignment_scores:
           Quality:  142.50            Length:    195
             Ratio:  1.197               Gaps:      7
 Percent Similarity:  61.026   Percent Identity: 21.538

alignment_block:
 US-09-303-518D-571 x Q9AKK5   ..

 Align seg 1/1  to: Q9AKK5  from: 1  to: 290

     319 GGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTT 368
         |||   |||:::::::::::::         :::|||:::   :::|||::
     104 GlyIleGluAsnIleLysLysLeu......GluGlyGlnProPheLeuLe 118

     369 CATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCC 418
         :::::::   |||:::::::::::|||::::::   :::   :::   :
     118 uPheSerGlyHisPheAlaAsnTrpAspIleSerLeuLysLeuLeuHisL 135

     419 AGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAA 468

         ::      |||      ::::::::::::|||:::      :::
     135 ysPheTyrPro...LysValAlaValIleTyrArgLysAlaAsnAsnPro 150

     469 GCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAAC 518
           ::::::|||:::::::::   :::|||:::      |||   :::
     151 TyrValAsnLysLeuValAsnGluSerArgAlaGlyAspLysLeuArgLe 167

     519 CGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCG 568
            |||   |||   :::|||   :::   ::::::::::|||::::::
     167 uIleProLysGlyProGluGlySerArgAlaLeuValArgAlaIleLysG 184

     569 CGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAA 618
         :::|||      ::::::|||   |||:::            :::
     184 luSerGluProIleValMetLeuValAspGln............LysMet 196

     619 GGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGAC 668
         :::   |||:::   :::   |||:::|||   ||||||   |||   ::
     197 AsnAspGlyIleGluValProPheLeuGlyHisProAlaMetThrAlaSe 213

     669 ACTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCT 718
         :      ||||||:::|||               |||      ::::::
     213 rAlaIleAlaLysIleAla......LeuGlnTyrLysTyrProIleIleP 228

     719 GCTGCGAACGCCTG...CCCGACGGACAAGGCTTCGTGTTGCACATCCGC 765
         |||:::   :::          |||:::   |||   :::   ::::::
     228 roCysGlnIleIleArgThrLysGlySerTyrPheLysValIleValHis 244

     766 CCC...GTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCC..... 807
         |||   ::::::   |||   :::|||::::::      |||:::
     245 ProGlnLeuLysPheGluGlnThrGlyAspAsnLysAlaAspCysTyrAs 261

     808 .......GCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTC 850
                    :::|||:::            |||:::::::::   |
     261 nIleMetLeuAsnIleAsnGlnMetLeuGlyGluTrpValLysGlnAsnP 278

     851 CGACGCAGTATCTGTTTATGTACAACCGCTATAAA 885
         ||:::|||:::::::::::::::::||||||:::|||
     278 roSerGlnTrpPheTrpPheHisAsnArgTrpLys 289

seq_name: sp_human:Q16824

seq_documentation_block:
ID   Q16824      PRELIMINARY;      PRT;   797 AA.
AC   Q16824;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PROFILAGGRIN (FRAGMENT).
GN   FLG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91064347; PubMed=2248957;
RA   Gan S.Q., McBride O.W., Idler W.W., Markova N., Steinert P.M.;
RT   "Organization, structure, and polymorphisms of the human profilaggrin
RT   gene [published erratum appears in Biochemistry 1991 Jun
RT   11;30(23):5814].";
RL   Biochemistry 29:9432-9440(1990).
DR   EMBL; M60502; AAA63248.1; -.
DR   InterPro; IPR003303; Filaggrin.
DR   PRINTS; PR00487; FILAGGRIN.
FT   NON_TER       1       1
SQ   SEQUENCE   797 AA;  85176 MW;  60E6184763BDA86B CRC64;


alignment_scores:
           Quality:  142.00            Length:    222
             Ratio:  1.193               Gaps:     10
 Percent Similarity:  53.604   Percent Identity: 24.775
```

```
alignment_block:
 US-09-303-518D-571 x Q16824   ..

 Align seg 1/1  to: Q16824  from: 1  to: 797

     162 CGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGC 211
         |||:::   |||||||||          |||:::        :::   ::
     266 ArgThrSerArgArgGln.........GlySerSerValSerGlnAspSe 279

     212 AGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTT 261
         :|||:::|||::::::         :::::::::   |||   |||:::: .
     279 rAspSerGlnGlyHis...SerGluAspSerGluArgArgSerGlySerA 295

     262 GCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGC 311
         ::   |||                  :::||||||:::::::::|||:::
     295 laSerArg.................AsnHisArgGlySerAlaGlnGlu 305

     312 GGTACACGGCTGGGAACACGT.......................GCAGC 337
            :::|||   |||:::|||                        ||
     306 GlnSerArgAspGlySerArgHisProArgSerHisHisGluAspArgAl 322

     338 AGGCTTTGGA............CAAGGGCGAAGGGCTGCTGTTCATCACG 375
         ||||   |||            :::::::|||:::::::::::::|||||||
     322 aGlyHisGlyAspSerAlaGluSerSerArgGlnSerGlyThrHisHisA 339

     376 CCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCT 425
         ||   :::   :::          :::   :::   |||:::   |||
     339 laGluAsnSerSerGlyGlyGlnAlaAlaSerSerHisGluGlnAlaArg 355

     426 TCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAG 475
         |||         :::||||||   :::      ::::::::::::|||
     356 SerSerAlaGlyGluArgHisGlySerHisTyrGlnGlnSerAlaAspSe 372

     476 ACAAAATCATGCAGGC.........GGGCAGGGTGCGCGGCAAAGGCAAA 516
          ::::::|||:::|||         ||||||::::::          ::::
     372 rSerArgHisSerGlyIleGlyHisGlyGlnAlaSerSerAlaValArgA 389

     517 ACCGCGCCCACCGGCATACAAGGGGTCAA......ACAAATCATCAAGGC 560
         ::   :::|||||||   :::   |||:::      ::::::::::::|||
     389 spSerGlyHisArgGlySerSerGlySerGlnAlaSerAspAsnGluGly 405

     561 CCTGCG..............................CGCGGGCGAGG 577
            :::                             |||      ||
     406 HisSerGluAspSerAspThrGlnSerValSerAlaHisArgGlnAlaGl 422

     578 CAACCATCATCCTGCCCGA.........CCACGTCCCTTCTCCGCAGGAA 618
         |:::||||||   :::::            |||      |||   ||||
     422 yArgHisHisGluSerHisGlnGluSerThrArgGlyArgSerArgGlyA 439

     619 GGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCAT... 665
         ||      |||   |||:::|||:::   |||:::      |||:::
     439 rgSerGlyArgSerGlySerPheLeuTyrGlnValSerThrHisGluGln 455

     666 ............................GACACTGGCGGCAAAATTGG 685
                                     :::||||||||||:::   ||
     456 SerGluSerAlaHisGlyArgAlaGlyProSerThrGlyGlyArgGlnGl 472

     686 CACACGTCAAAGGCGT 701
         |:::|||:::::::::
     472 ySerArgHisGluGln 477

seq_name: sp_human:Q9H4U3

seq_documentation_block:
ID   Q9H4U3      PRELIMINARY;      PRT;   798 AA.
AC   Q9H4U3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DJ14N1.1.2 (PROFILAGGRIN 3' END) (FRAGMENT).
GN   FLG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Laird G.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356504; CAC13171.1; -.
DR   InterPro; IPR003303; Filaggrin.
DR   PRINTS; PR00487; FILAGGRIN.
FT   NON_TER       1      1
SQ   SEQUENCE   798 AA;  84773 MW;  F923DDA8D1290805 CRC64;


alignment_scores:
           Quality:  142.00              Length:    276
             Ratio:  1.164                 Gaps:     10
 Percent Similarity:  44.203   Percent Identity: 21.377

alignment_block:
 US-09-303-518D-571 x Q9H4U3   ..

 Align seg 1/1  to: Q9H4U3  from: 1  to: 798

     111 GCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACC 160
         :::|||      :::|||   ||||||
     232 SerGlyAspGlySerArgHisSerGly..................... 240

     161 GCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACG 210
              :::|||:::::::::   ::::::   :::   :::   ||||||:
     241 ....SerHisHisHisGluAlaSerSerTrpAlaAspSerSerArgHisS 256

     211 CA...GACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGA 257
         ::   :::|||||||::::::   :::|||
     256 erLeuSerGlyGlnGlyGlnSerAlaGly..................... 265

     258 ACTTGCCCCCGCGTTTTTCAAAAAACCGGA..................... 287
                |||||||:::   ::::::   |||
     266 ......ProArgThrSerArgAsnGlnGlySerSerValSerGlnAspSe 280

     287 ........................................................ 287

     280 rAspSerGlnGlyHisSerGluAspSerGluArgTrpSerGlySerAlaS 297

     288 .....AGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACGT 332
              :::|||||:::::::::|||:::   :::|||   |||:::|||
     297 erArgAsnHisArgGlySerAlaGlnGluGlnSerArgAspGlySerArg 313

     333 ........................GCAGCAGGCTTTGGA........... 347
                                    ||||||   |||
     314 HisProThrSerHisHisGluAspArgAlaGlyHisGlyHisSerAlaGl 330

     348 .CAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGAT 396
          :::::::|||:::::::::::::|||||||||   :::   :::
     330 uSerSerArgGlnSerGlyThrHisHisAlaGluAsnSerSerGlyGlyG 347

     397 TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCAT 446
           :::   :::   |||:::   |||   |||          :::||||||
     347 lnAlaAlaSerSerHisGluGlnAlaArgSerSerAlaGlyGluArgHis 363

     447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGC..... 491
         :::       :::::::::::|||   ::::::|||:::|||
     364 GlySerHisHisGlnGlnSerAlaAspSerSerArgHisSerGlyIleGl 380

     492 ....GGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAA 537
             ||||||::::::          ::::::   :::|||||||   :::
     380 yHisGlyGlnAlaSerSerAlaValArgAspSerGlyHisArgGlySerS 397

     538 GGGGT.......................................... 542
           |||
     397 erGlySerGlnAlaSerAspSerGluGlyHisSerGluAspSerAspThr 413

     543 ................................CAAACAAATCATCA 556
                                         |||   :::||||||
```

```
     414 GlnSerValSerAlaHisGlyGlnAlaGlyProHisGlnGlnSerHisGl 430

     557 AGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCT 606
         |      :::||||||||||
     430 nGluSerThrArgGlyArg.............................. 436

     607 TCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGC 656
         ||||||||||||      |||   |||:::|||:::   |||:::
     437 ..SerAlaGlyArgSerGlyArgSerGlySerPheLeuTyrGlnValSer 452

     657 ATACACCAT.................................GACACTGG 673
         |||:::                                 :::|||||
     453 ThrHisGluGlnSerGluSerAlaHisGlyArgAlaGlyProSerThrGl 469

     674 CGGCAAAATTGGCACACGTCAAAGGCGT 701
         ||||:::   |||:::|||:::::::::
     469 yGlyArgGlnGlySerArgHisGluGln 478

seq_name: sp_bacteria:Q52747

seq_documentation_block:
ID   Q52747      PRELIMINARY;      PRT;   680 AA.
AC   Q52747;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ENDO-GLUCANASE.
OS   Ruminococcus flavefaciens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Ruminococcus.
OX   NCBI_TaxID=1265;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=186;
RA   Huang C.M., Asmundson R.V., Yu P.L.;
RT   "Nucleotide sequence of a cellulase gene complex from Ruminococcus
RT   flavefaciens strain 186 coding for multi-cellulase activities.";
RL   Submitted (SEP-1990) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X56082; CAA39559.1; -.
SQ   SEQUENCE   680 AA;  75503 MW;  5A589F334846DD6F CRC64;


alignment_scores:
            Quality:  141.50            Length:    316
              Ratio:  1.004               Gaps:     19
 Percent Similarity:  44.620   Percent Identity: 27.848

alignment_block:
 US-09-303-518D-571 x Q52747   ..

 Align seg 1/1  to: Q52747  from: 1  to: 680

      29 CCCCTTTGCGAACCGCCATGCACATCCTGTTGACCG...CCCTGCTCAAA 75
         |||           |||   |||:::      |||   |||   |||
      27 ProThrProProSerProSerThrGluSerSerProThrProSerSerPr 43

      76 TGCCTCTCCC..............TGCTGTCGC.........TTTCCTG 101
          |||||||||                |||||             |||:
      43 oAlaSerProArgSerProThrValArgCysArgGluAlaValAlaProA 60

     102 TCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAA 151
         ::||| ||||||         |||||||||:::|||           ::
      60 laCys.HisAla......AspAlaArgThrAlaGly............Ar 70

     152 AGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAAC 201
         :||||||                                    :::|
      70 gGlyArg.....................................AspP 74

     202 CCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGG 251
         ||||||||   :::   ::::::|||:::   |||:::   :::   |||
      74 roArgHis...GlnCysAspAlaArgLeuArgGlyArgHisArgProArg 89

     252 TTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAA 301
         :::|||   :::           :::      |||   |||:::::::::
      90 LeuGlyHisAlaValValArgAspAspGlyLeuGlyProHisGlnArgAs 106

     302 TGTTCAAAGCGGTACACGGCTGGGA......ACACGTGCAGCAGGCTTTG 345
         :         |||:::      |||      ::::::   |||   :::
     106 pCysValArgGlyAlaCysHisGlyProSerValHisArgAlaAspTrpH 123

     346 GACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGA 395
           |||   |||||||||||:::::::
     123 isGlnProArgArgAlaAlaLeuGlu.....................Leu 132

     396 TTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCA 445
           |||:::   :::|||      ::::::   ||||||      |||
     133 ValGlyGlnLeuValHisProGlnSerGlyLeuValProValValArgGl 149

     446 TGTACAAGCC..................GCCGAAAATCAAAGC....... 470
          |||   |||                      ::::::||||||
     149 yValProAlaProProProArgAspArgArgArgAspGlnSerArgProA 166

     471 ...............................GATAGACAAAAT...... 482
                                         ||||||::::::
     166 rgArgAlaLeuValGlnLeuProLeuProArgAspArgGluGlnProGly 182

     483 ..............................CATGCAGGCGGGCAGGGTGC 502
                                      |||:::||||||:::|||::
     183 GlyHisGlnProHisAspAlaAspAlaValHisGlyGlyGlyArgGlyVa 199

     503 GCGG...CAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAA 549
         :|||   |||   |||:::|||      |||      |||   :::   :
     199 lArgGlyGlnLeuGlnSerArgArgProArgAspArgArgValArgAspH 216

     550 ATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCA 599
         ::::::::||||||:::   ||||||::         |||   |||
     216 isArgAspGlyProSerAspGlyArgSerPheAspProProAspArgGln 232

     600 CGTCCCTTC......TCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTT 643
         |||            :::   |||||||||:::|||:::
     233 ArgAlaValArgArgGlnArgGlyArgArgHisArgLeu........... 245

     644 TCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTC 693
                                              :::      |||
     246 ................................ValAspArgArgV 250

     694 AAAGG...CGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGG 740
           ||   ||||||:::|||||||         |||:::      |||
     250 alArgProArgGluGlnProVal......ProArgValArgGlnArgGly 264

     741 ACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGA........... 779
               :::|||:::   :::      |||   ||||||
     265 ProGlyValArgAlaHisGluHisArgArgArgArgGlyProLeuAlaGl 281

     780 .......ATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAA 818
                 :::||||||   |||:::   :::   |||   ||||||
     281 yAspLeuValGluArgAlaGlnAlaAlaHisProArgSerValGln 296

seq_name: sp_bacteriap:Q9CNR4

seq_documentation_block:
ID   Q9CNR4      PRELIMINARY;      PRT;   318 AA.
AC   Q9CNR4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MSBB.
GN   MSBB OR PM0362.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
```

```
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR   EMBL; AE006072; AAK02446.1; -.
KW   Complete proteome.
SQ   SEQUENCE   318 AA;  36768 MW;  F91DDAB6D9C46375 CRC64;


alignment_scores:
           Quality:  141.00            Length:    268
             Ratio:  0.916               Gaps:     10
 Percent Similarity:  57.463   Percent Identity: 21.269

alignment_block:
 US-09-303-518D-571 x Q9CNR4   ..

 Align seg 1/1  to: Q9CNR4  from: 1  to: 318

     112 CTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCG 161
         |||::::::   :::|||   ::::::           |||:::::||
      53 LeuAlaSerTyrIleGlyLysIleValAlaArgLysAlaLysLysGlnAr 69

     162 CGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAAC..........  201
         |   |||          |||::::::   :::      :::
      69 gHisArgAlaLysIleAsnLeuGlnTyrCysPheProHisTrpThrGluA 86

     202 ..CCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGC 249
                 ||||||::::::   :::|||:::   :::::::::
      86 laGlnArgGluGlnThrIleGluGluMetPheValIleValThrGlnVal 102

     250 GGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAAC 299
         |||   ::::::         :::::::::   :::   ::::::
     103 MetLeuGlyIleGlyGluIleAlaIleArgSerLysArgHisLeuGlnAr 119

     300 AATGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACA 349
             ::::::     |||   ||||||:::::::::|||   ::::
     119 gArg...SerCysPheThrGlyLeuGluHisIleHisArgAlaArgGluG 135

     350 AGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTG 399
         ::|||:::::::::::::::::::::::||||||      :::   |||
     135 lnGlyLysAsnIleIleLeuLeuValProHisAlaTrpAlaIleAspAla 151

     400 GGCGGACGCTACATC...AGCCAGCAGCTTCCGTTCCACCTGACCGCCAT 446
         :::|||      :::   ::::::   :::|||      :::|||:::||
     152 SerGlyIleIleLeuHisThrHisGlyMetPro......MetThrSerMe 166

     447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCA 496
         ||||:::|||   :::         :::|||   :::         :::|
     166 tTyrAsnProHisArgAsnProLeuValAspTrpLeuTrpThrPheAlaA 183

     497 GGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAA 546
         ||   |||      ||||||                :::|||:::|||
     183 rgGlnArgPheGlyGlyLysMetHisAlaArgGln...AsnGlyIleLys 198

     547 CAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGA 596
            :::::::::   ::::::   |||:::            ||||||||
     199 ProPheLeuAsnHisValLysGlnGlyAspMetGlyTyrTyrLeuProAs 215

     597 CCACGTCCCTTCTCCGCAGGAAGGCGGC.........GGCGTGTGGGCGG 637
         |                ::::::   |||         :::|||::::::|
     215 p..............GluAspTyrGlyAlaGluLeuSerValPheValA 227

     638 ATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCA 687
         ||||||||:::            |||:::            |||:::|||
     227 spPhePheAlaThrTyrLysAlaThrLeuProGlyIleAsnLysIleAla 243

     688 CACGTC...AAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCC 734
         ::::::   |||::::::         :::|||         :::
     244 ArgLeuAlaLysAlaAlaValIleProMetPheProArgTyrAsnAlaLe 260

     735 CGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGA 784
          :::|||:::   :::   ::::::|||:::|||::::::      |||:
     260 uSerGlyLys...TyrGluIleGluIleHisProAlaMet...ThrLeuS 275

     785 ACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAATACCGAATAT 834
         ::   :::      :::   |||   ::::::|||::::::   |||
     275 erAspAspProGluGlnAlaAlaArgAlaMetAsnGlnGluIleGluSer 291

     835 TGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTATAA 884
         ::::::         |||:::||||||::::::::   :::      ||
     292 PheValThrAlaThrProAlaGlnTyrValTrpIleLeuArgLeuLeuLy 308

     885 AACG 888
         ||||
     308 sThr 309

seq_name: sp_bacteria:Q51415

seq_documentation_block:
ID   Q51415      PRELIMINARY;      PRT;   684 AA.
AC   Q51415;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ALGINATE LYASE.
GN   ALGY.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A22 ALG+;
RA   Schmitt Andrieu L., Hulen C.;
RT   "Cloning and sequencing of a new alginate lyase, algY, from mucoid
RT   strains of Pseudomonas aeruginosa, and its expression in E. coli.";
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z54213; CAA90940.1; -.
KW   Lyase.
SQ   SEQUENCE   684 AA;  74958 MW;  CBCCE638F5675D7A CRC64;


alignment_scores:
           Quality:  140.50            Length:    270
             Ratio:  1.115               Gaps:     10
 Percent Similarity:  46.667   Percent Identity: 26.296

alignment_block:
 US-09-303-518D-571 x Q51415   ..

 Align seg 1/1  to: Q51415  from: 1  to: 684

      42 CGCCATGCACATCCTGTTGAC...CGCCCTGCTCAAATGCCTCTCCCTGC 88
         |||:::   ||||||   :::      |||:::            |||::
     304 ArgArgArgHisProPheSerAlaThrProSerAlaThrArgSerProTh 320

      89 TGTCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCG 138
         :   :::      |||:::   |||::::::||||||      :::|||
     320 rSerSer......SerGlyLeuAlaAlaArgProAlaGlyArgGlyGlyG 335

     139 TTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCA 188
           ||||||   :::::::::||||||::::::|||:::         :::
     335 lyLeuProArgGlnAlaHisProArgGlyGluArgGlnGlyAlaLeuSer 351

     189 GGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTTTGCGGAAA 238
         ::::::               |||            ||||||   |||
     352 AlaAlaAlaThrAlaValSerAlaAlaValLeuArgArgPheLeuGly.. 367

     239 CGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAA 288
                                                 ||||
     368 .....................................GlyA 369

     289 GACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACA......CGT 332
         ||   |||::::::   ::::::|||   ||||||            |||
     369 rgProArgThrGlyArgArgThrGlyGlyArgLeuProAspLeuGlyArg 385
```

```
     333 GCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACA 382
         |||:::|||   |||   |||::::::               |||:::
     386 AlaValGlyGlyGlyGlyGlyGluAsnArgArgArgAlaProAlaSerGl 402

     383 TCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTC 432
                   :::|||::::::   |||   ||||||::::::   |
     402 ySerProTrpAlaTyrGlyGluValArgHisProProAlaCysAspArgP 419

     433 CACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAAT 482
         ||   :::|||:::   :::   |||::::::
     419 roArgAsnArgArgArgSerLeuAlaArgArg..................... 429

     483 CATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCA 532
                 ||||||:::   |||   :::   :::|||   ||||||::
     430 .........GlyGlnAlaAspArgThrHisLeuArgArgAsnHisArgAr 443

     533 T.....................................ACAAGGG 540
         :                                     |||:::|
     443 gArgAlaAlaValLeuArgProPheArgSerGlyArgProAlaThrHisG 460

     541 GTCAAACAAATCATCAA.................GGCCCTGCGCGCGGG 572
         ||::::::   ::::::                |||||:::
     460 lyArgAlaAlaArgArgProProArgSerProGlyAspProAlaGluPro 476

     573 CGAGGCAACCATCATCCT...GCCCGACCACGTCCCTTCTCCGCAGGAAG 619
            |||:::   :::|||   ||||||   |||         ||||||||
     477 LeuGlyArgGlyArgProGlyAlaArgArgArgArgHisArgAlaGlyAr 493

     620 GCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACA 669
         ||||      :::|||||||
     493 gArgSerAlaAlaGlyGly.............................. 499

     670 CTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTG 719
           |||:::            |||:::   |||::::::   :::   |||
     500 ..GlyAlaHisArgArgIleArgArgAlaArgHisArgGlnLeuHisLeu 515

     720 CTGCGA.....................ACGCCTGCCCGACGGACAAGGCT 748
         ||||||                   |||   |||||||||   ||
     516 LeuArgLeuSerAlaProGlyArgGlyLeuProLeuArgArgThrGlyLe 532

     749 TCGTGTTGCA 758
         |   :::|||
     532 uProAlaAla 535

seq_name: sp_bacteria:Q9F6E3

seq_documentation_block:
ID   Q9F6E3      PRELIMINARY;      PRT;   320 AA.
AC   Q9F6E3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PKSA PUTATIVE O-METHYLTRANSFERASE.
OS   Streptomyces collinus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=42684;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM2012;
RA   Martin R., Bailey J.E., Minas W.;
RT   "Putative spore pigment polyketide synthase cluster from Streptomyces
RT   collinus DSM2012.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF293354; AAG26883.1; -.
KW   Transferase; Methyltransferase.
SQ   SEQUENCE   320 AA;  35453 MW;  AEDF4519923A166F CRC64;


alignment_scores:
           Quality:  138.50            Length:    324
             Ratio:  1.011               Gaps:     16
 Percent Similarity:  42.284   Percent Identity: 26.235
```

```
alignment_block:
 US-09-303-518D-571 x Q9F6E3   ..

 Align seg 1/1  to: Q9F6E3  from: 1  to: 320

      27 TCCCCCTTTGCGAACCGCCATGCACATCCTGTT.........GACCGCCC 67
         ||||||         |||:::   |||   :::         |||
      44 SerProValLeuProArgArgLeuHisGlyThrProGlyArgAspValAr 60

      68 TGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGA 117
          |||:::      |||||||||:::|||                   |||:
      60 gAlaHisArgArgLeuProAlaAlaAla..............ArgGlyA 72

     118 AACCGGCTCGGACATCT.............................. 134
         ::|||   :::   :::
      72 rgProProGlnProAlaArgHisHisAlaValValHisArgAlaValAsp 88

     135 ...GGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATA 181
            |||   |||         ||||||      |||||||||   |||
      89 LeuGlyArgLeuAlaGluAlaGlyArgGlyGlyAlaHisArgProGlnAr 105

     182 TGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTTT 231

     105 g.............................................. 105

     232 GCGGAAACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAA 281
         |||:::|||   :::|||   |||:::   |||
     106 ..GlyGlyGlyProValArgGlnGlyValLeuProLeu.......... 117

     282 ACCGGAAGACATCGAAACAATGTTCAAAGCGGTACACGGCTGGGAACACG 331
                   :::|||::::::         |||   ||||||      ||
     118 ......ProGlnArgGlyArgProGlyValGlyGlyArgLeuGlnProAr 132

     332 T............GCAGCAGGCTTTGGACAAGGGCGAAGGGCT....... 362
         |                  :::   |||   ||||||||
     132 gHisAspAspLeuGlnArgAlaValGlyAlaGlyArgArgArgValProA 149

     363 ........GCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGG 404
                 ::::::|||   ::::::::::|||:::   |||   |||
     149 rgProHisGlyMetHisLeuGlyGlyArgArgArgArgArgProGlyAla 165

     405 ACGCTACATCAGCCAGCAGCTTCCGTTCCA...CCTGACCGCCATGTACA 451
                   :::||||||   :::||||||   |||   |||
     166 ArgArgGlyGluProAlaArgGlnValProGlyProAlaArgAsp***Al 182

     452 AGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTG 501
                         :::   :::::::::         ::::::|||
     182 aArgProProProGlyGlyGlyGluArgArgProGluThrAlaHisGlyA 199

     502 CGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAAT 551
           |||   |||      ||||||:::|||      |||         :::
     199 rgArgProArgGlyProArgAlaArgArgAlaArgArgArgProAspGly 215

     552 CATCAAGGC.............................CCTGCGCGCGG 571
         :::   |||                              |||:::   ||
     216 ArgProGlyGluGlyArgArgLeuArgAspGlnGluHisProGlyValGl 232

     572 GCGAGGCAACCATCATCCT.........GCCCGACCACGTCCCTTCTCCG 612
         ||||   :::|||||||||            |||||||||
     232 yArgArgGluHisHisProHisAlaAlaGlnArgProArg.......... 245

     613 CAGGAAGGCGGCGGCGT......GTGGGCGGATTTTTTCGGCAAACCTGC 656
           |||||||||   |||         ||||||
     246 ..GlyArgArgProArgAspAlaArgGlyGly.................. 255

     657 ATACACCATGACACTGGCGGCAAAATTGGC......ACACGTCAAAGGCG 700
           |||:::::::::      :::::::::      :::   |||:::||
     256 ...HisArgGlu***ArgArgArgLeuAlaLeuAspAlaValGlnHisAr 271

     701 TGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTC 750
         |:::::|||                  :::   |||   :::   |||
     271 gAspGlyPro..............AlaAlaAlaProGlnArgArgArgG 283
```

```
     751 GTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAAGCCCA 800
         |||||||||   :::      |||      ::::::     |||
     283 lyGlnAlaHisProAlaGluHisGlyGlyProAlaHisArgGlyArgPro 299

     801 CGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTC 850
         ||||||   ||||||                          |||:::||
     300 GlyHisArgArgHisGlnPro.............GlyGlnProLeuSe 311

     851 CGACGCAGTATCTGTTTATGTA 872
         |     |||   :::::::::
     311 rAlaArgValArgLeuHisIle 318

seq_name: sp_bacteriap:Q9A6A1

seq_documentation_block:
ID   Q9A6A1      PRELIMINARY;      PRT;   424 AA.
AC   Q9A6A1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN CC2193.
GN   CC2193.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=69394;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005891; AAK24164.1; -.
DR   TIGR; CC2193; -.
DR   InterPro; IPR002048; EF-hand.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   424 AA;  47293 MW;  4E1FE0E40A22F933 CRC64;


alignment_scores:
           Quality:  138.00            Length:    305
             Ratio:  0.986               Gaps:     18
 Percent Similarity:  45.902   Percent Identity: 27.213

alignment_block:
 US-09-303-518D-571 x Q9A6A1   ..

 Align seg 1/1  to: Q9A6A1  from: 1  to: 424

      30 CCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCCTGCTCAAATGCC 79
         ||||||::::::|||                ||||||   :::
       5 ProPheValSerArg.................ArgProAspArgArgAr 15

      80 TCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCT.... 125
          ||||||               |||:::|||   :::|||:::
      15 gAlaProAla.........ArgArgAlaArgAlaArgArgProGlyProP 29

     126 ...................CGGACATCTGGCGTTTTACCTTTTAAAGGA 155
                            |||:::      |||   |||   :::|||
      29 heAlaLeuSerValAlaAspArgValArgCysValArgPro...GlnGly 44

     156 AGACCGCGCGCG.........CATCGTCGC...CAATATGCGGCAGGCGG 193
         |||      |||         ||||||:::      :::|||   |||||
      45 ArgHisGlyAlaGlyGlyLeuHisArgGluLeuLeuHisAlaArgGlyGl 61

     194 GTTTGAACCCCGA.............................CACG 210
         |   |||   |||                               |||
      61 yAspGluIleArgArgLeuArgArgAspArgAlaArgTrpArgProHisA 78

     211 CAGACGGT.........CAAAGCCGTTTTTGCGGAAACGGCAAAATGCGG 251
              |||         :::::::|||   |||:::::::|||
      78 rgProGlySerGlyLeuHisAlaArgAlaCysAlaGlyGlyGly...... 92

     252 TTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAA 301
                                 ::::::   :::||||||   :::::
      93 .....................AspAspArgGluAlaArgHisLeuSerAs 102

     302 TGTTCAAAGCGGTACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAG 351
         :   :::   ::::::|||                ::::|||:::|||||||
     102 pHisHisArgSerAlaArgArgProArgGlyGlnGlyGlyLeuGlyGlnG 119

     352 GGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGG 401
         |||||   :::::::::               |||:::|||:::   :::
     119 lyArgValGlyValLeu..............ArgArgLeuGlnGlyAla 130

     402 CGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACA 451
                 |||:::   |||          |||      |||      ::
     131 GlyArgArgHisArgHisAlaGluProAspProGlyArgArgLeuCysAs 147

     452 AGCCGCCGAAAATCAA..............AGCGATAGACAAAATCATG 486
         :|||   :::   |||               :::   ||||||:::   :
     147 pAlaArgArgAlaGlnArgCysGlyLeuGlyAlaAlaArgGlnGlySerG 164

     487 CA......GGCGGGCAGGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGG 530
         ::      |||   :::   :::::::|||   :::::::|||      |||
     164 lyValProGlyArgAspProSerGlnGlnAspArgHisArgProGlyArg 180

     531 CATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGCGGGCGAGGCAA 580
         ::::::||||||   :::::::::|||      ||||||||||||
     181 ArgValArgGlyAlaValArgArgGln......AlaArgGlyArgProAl 195

     581 CCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTG 630
          :::|||||||||                          ||||||||||||
     195 aAsnHisProAla.....................ArgArgArgArgP 204

     631 TGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAAA 680
                                  :::::::::      ||||||
     204 ro.....................LeuArgArgHisGlnGlyGlyAla 212

     681 ATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCC 730
               ::::::::::||||||::::::|||          ||||||   ||
     213 GlnAspAlaGluGluArgArgGlnArgPro.........LeuArgArgPr 226

     731 TGCCCGA.................................CGGACAA 744
         |:::|||                                    |||   |
     226 oGlyArgAlaValGlyArgLeuProAspProValAlaLeuProArgArgA 243

     745 GGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAA 794
         ||   |||            |||||||||   |||   |||   :::
     243 rgAlaArg............ProArgProAlaGlyProGluLeuHisLeu 255

     795 AGCCCACGATGCCGC 809
             |||      |||
     256 ArgProAspLeuArg 260

seq_name: sp_bacteriap:Q9KVD4

seq_documentation_block:
ID   Q9KVD4      PRELIMINARY;      PRT;   325 AA.
AC   Q9KVD4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIPID A BIOSYNTHESIS (KDO)2-(LAUROYL)-LIPID IVA ACYLTRANSFERASE.
GN   VC0212.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
```

```
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004111; AAF93388.1; -.
DR   TIGR; VC0212; -.
KW   Complete proteome.
SQ   SEQUENCE   325 AA;  37215 MW;  3F949ED94B5C5F92 CRC64;


alignment_scores:
            Quality:  137.50              Length:    193
              Ratio:  1.239                 Gaps:      6
 Percent Similarity:  57.513   Percent Identity: 24.870

alignment_block:
 US-09-303-518D-571 x Q9KVD4   ..

 Align seg 1/1  to: Q9KVD4  from: 1  to: 325

     319 GGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTT 368
         |||   |||::::::         ||||||:::||||||   :::::::::
     122 GlyGluGluAsnLeuLeuProLeuLeuAspSerGlyGluArgValIleLe 138

     369 CATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCC 418
         :::::::||||||      :::   |||   ::::::      ::::::
     138 uLeuValProHisSerTrpAlaIleAspTyrAlaAlaValMetLeuAlaA 155

     419 AGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAA 468
           :::       :::   ::::::   :::   ||||||   :::
     155 laArg...GlyTyrLysValAlaAsnIleMetLysProGlnArgAsnPro 170

     469 GCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAAC 518
               |||   :::|||::::::   |||::::::      |||:::
     171 IleAlaAspTrpLeuMetHisValGlnArgMetGlnTyrGlyGlyArgIl 187

     519 CGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCG 568
               |||       :::|||:::|||   ::::::::::::::::::
     187 ePhe...ThrArgGluSerGlyIleLysProPheLeuArgSerIleGlnS 203

     569 CGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAA 618
         ::|||   :::         :::||||||:::      :::||||||
     203 erGlyTyrValGlyTyrTrpValProAspGluAspHisGlyProGln... 218

     619 GGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGAC 668
           :::::::|||::::::   ||||||:::           |||:::
     219 ...AsnSerValPheValProPhePheAlaThrGluLysAlaThrLeuLy 234

     669 ACTGGCGGCAAAATTGGCACACGTC......AAAGGCGTGAAAACCCTGT 712
               :::|||:::|||   :::           |||      |||:
     234 sGlyPheGlyLysMetAlaLysLeuCysLysAlaHisValValProLeuM 251

     713 TTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATC 762
         ::   |||   :::         |||:::      :::   :::||||||
     251 etSerCysTyrAsn......SerAspSerGlyArgTyrGluValHisIle 265

     763 CGCCCC...GTCCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGC 809
           |||   :::|||:::      :::|||::::::      ||||||
     266 LeuProAlaLeuGlnAsnPheProThrGlyAspGluGluAlaAspAlaLe 282

     810 CGTGTTCAACCGCAATACCGAATATTGGATACGCCGTTTTCCGACGCAGT 859
          ::::::||||||      |||      :::         |||   ||||
```

```
     282 uAlaMetAsnArgAlaIleGluAlaLeuValThrProGlnProGluGlnT 299

     860 ATCTGTTTATGTACAACCGCTATAAAACG 888
         ||::::::      :::      ||||||
     299 yrMetTrpAsnLeuSerLeuLeuLysThr 308

seq_name: sp_bacteriap:Q9K217

seq_documentation_block:
ID   Q9K217      PRELIMINARY;      PRT;   462 AA.
AC   Q9K217;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE, PUTATIVE.
GN   CP0676.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
DR   EMBL; AE002225; AAF38487.1; -.
DR   TIGR; CP0676; -.
DR   InterPro; IPR000504; RRM.
DR   PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW   Transferase; Acyltransferase.
SQ   SEQUENCE   462 AA;  52624 MW;  C31CD0C9B01E13DE CRC64;


alignment_scores:
            Quality:  137.00              Length:    338
              Ratio:  0.867                 Gaps:      15
 Percent Similarity:  46.746   Percent Identity: 23.964

alignment_block:
 US-09-303-518D-571 x Q9K217   ..

 Align seg 1/1  to: Q9K217  from: 1  to: 462

      52 ATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTG 101
         ||||||   :::   |||         :::|||   :::         ||
      12 IleLeuGluAlaProLeuTyrTyrLeuValSerGlyIleIleAlaLeuCy 28

     102 TCTGCACACG.....................CTGGGAAACCGGCTCGGAC 130
         |   ||||||                    ||||||:::   :::|||
      28 sArgHisThrProArgSerPheLeuThrGlyLeuGlyLysGlyPheGlyP 45

     131 ATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAAT 180
           ||||||||||||:::::::::::::   |||         :::::|||
      45 heLeuAlaPheTyrIleIleSerAspTyrArgLysThrAlaLeuThrAsn 61

     181 ATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTT 230
         :::      |||                          :::   :::||
      62 LeuAlaLeuAla.....................PheProGluLysThrPh 71

     231 TGCGGAAACGGCAAAATGCGGT........................... 252
         |   |||      |||   :::
      71 eAspGluArgTyrLysIleAlaArgGlnSerLeuGlnHisLeuIleIleT 88

     253 .....TTGGAACTTGCC.............................. 264
               ||||||||
      88 hrLeuLeuGluLeuLeuAlaIleGluGlnLeuValGlyAsnIleAspLys 104
```

```
 265 ..............................CCCGCGTTTTTCAAAAAACC 284
                                  |||      |||::::::
 105 LeuIleThrIleValThrSerSerArgAsnProLysGlyPheSerSerGl 121

 285 G..............GAAGACATCGAAACAATGTTCAAAGCGGTACACG 319
                    ||||||:::|||      ||||||
 121 uGluValIleSerAsnGluAspLeuGluGluThrPheLys..... 134

 320 GCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTC 369
                          |||::::::   :::||||||::::::
 135 ...................AsnLeuGlnGluLysGlnGlyLeuIleLeu 144

 370 ATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCA 419
     :::      |||   ::::::::::::|||         ||||||:::::
 145 PheCysGlyHisGlnAlaAsnTrpGluLeuProPheLeuTyrIleThrLy 161

 420 GCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAG 469
     ::::   |||            |||:::   |||      |||   :::
 161 sAsnTyrProGly......IleAlaPheAlaLysAlaIleLysAsnGlnA 176

 470 CGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACC 519
       :::::::|||   :::   |||  |||          |||||||||
 176 rgLeuSerLysLysIlePheAlaLeuArgGluValPheLysGlyLysIle 192

 520 GCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGC 569
     :::|||         :::|||::::::|||   |||:::|||||:::
 193 ValProProLys...AsnGlyIleGlnGlnGlyIleGluAlaLeuAsnGl 208

 570 GGGCGAGGCAACCATCATCCTGCCCGACCACGTC.............. 603
      |||:::   :::   |||:::   |||::::::
 208 nGlyLysLeuValGlyIleValGlyAspGlnAlaLeuLeuMetSerSerT 225

 604 ........CCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTC 645
             |||                                  :::|||
 225 yrThrTyrPro..............................LeuPhe 230

 646 GGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAA 695
     |||:::||||||:::|||   |||      |||   |||||||:::
 231 GlySerProAlaPheThrThrThrSerProAlaLeuLeuAlaTyrLysTh 247

 696 AGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAG 745
      |||      ::::::          :::|||      :::::::::|
 247 rGlyProValIleAlaValAsnValSerArg......GlnAlaLysG 262

 746 GCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAA 795
     |||||      ::::::   |||   :::::::::|||   :::|||||
 262 lyPhe......GluValIlePro...SerAlaLysLeuTyrAlaAsnLys 275

 796 GCC............CACGATGCCGCCGTGTTCAACCGCAAT........ 825
     :::            :::::::::|||::::::      :::
 276 SerLeuProMetLysGluSerValAlaIleLeuMetAspGlnMetMetGl 292

 826 ....ACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGT 871
             |||      |||         |||   |||:::::::::::::
 292 yPheLeuGluLysGlyIleAlaSerGlnProGluGlnTrpMetTrpIleH 309

 872 ACAACCGCTATAAA 885
     :::::|||:::|||
 309 isLysArgTrpLys 313
```

```
seq_name: sp_bacteriap:Q9Z983

seq_documentation_block:
ID   Q9Z983      PRELIMINARY;      PRT;   467 AA.
AC   Q9Z983;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ACYLTRANSFERASE.
GN   HTRB OR CPN0098.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
DR   EMBL; AE001596; AAD18251.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   467 AA;  53193 MW;  D3C7C284E9220DD0 CRC64;


alignment_scores:
           Quality:  137.00              Length:    338
             Ratio:  0.867                Gaps:     15
 Percent Similarity:  46.746   Percent Identity: 23.964

alignment_block:
 US-09-303-518D-571 x Q9Z983   ..

 Align seg 1/1  to: Q9Z983  from: 1  to: 467

  52 ATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTG 101
     ||||||   :::   |||         :::|||   :::         ||
  12 IleLeuGluAlaProLeuTyrTyrLeuValSerGlyIleIleAlaLeuCy 28

 102 TCTGCACACG.....................CTGGGAAACCGGCTCGGAC 130
     |   ||||||                     ||||||:::   :::|||
  28 sArgHisThrProArgSerPheLeuThrGlyLeuGlyLysGlyPheGlyP 45

 131 ATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAAT 180
       ||||||||||||:::::::::::::   |||         ::::::|||
  45 heLeuAlaPheTyrIleIleSerAspTyrArgLysThrAlaLeuThrAsn 61

 181 ATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTTT 230
     :::      |||                          :::   :::||
  62 LeuAlaLeuAla.....................PheProGluLysThrPh 71

 231 TGCGGAAACGGCAAAATGCGGT........................... 252
     |   |||      |||   :::
  71 eAspGluArgTyrLysIleAlaArgGlnSerLeuGlnHisLeuIleIleT 88

 253 .....TTGGAACTTGCC.............................. 264
          |||||||||
  88 hrLeuLeuGluLeuLeuAlaIleGluGlnLeuValGlyAsnIleAspLys 104

 265 ..............................CCCGCGTTTTTCAAAAAACC 284
                                  |||      |||::::::
 105 LeuIleThrIleValThrSerSerArgAsnProLysGlyPheSerSerGl 121

 285 G..............GAAGACATCGAAACAATGTTCAAAGCGGTACACG 319
                    ||||||:::|||      ||||||
 121 uGluValIleSerAsnGluAspLeuGluGluThrPheLys..... 134

 320 GCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTC 369
                          |||::::::   :::||||||::::::
 135 ...................AsnLeuGlnGluLysGlnGlyLeuIleLeu 144

 370 ATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCA 419
     :::      |||   ::::::::::::|||         ||||||:::::
 145 PheCysGlyHisGlnAlaAsnTrpGluLeuProPheLeuTyrIleThrLy 161

 420 GCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAG 469
     ::::   |||            |||:::   |||      |||   :::
 161 sAsnTyrProGly......IleAlaPheAlaLysAlaIleLysAsnGlnA 176

 470 CGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACC 519
       :::::::|||   :::   |||  |||          |||||||||
 176 rgLeuSerLysLysIlePheAlaLeuArgGluValPheLysGlyLysIle 192

 520 GCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGCGCGC 569
     :::|||         :::|||::::::|||   |||:::|||||:::
```

```
     193 ValProProLys...AsnGlyIleGlnGlnGlyIleGluAlaLeuAsnGl 208

     570 GGGCGAGGCAACCATCATCCTGCCCGACCACGTC.............:... 603
         |||:::   :::   |||:::   |||::::::
     208 nGlyLysLeuValGlyIleValGlyAspGlnAlaLeuLeuMetSerSerT 225

     604 ........CCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTC 645
                 |||                                 :::|||
     225 yrThrTyrPro................................LeuPhe 230

     646 GGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAA 695
         |||:::||||||:::|||   |||      |||   ||||||:::
     231 GlySerProAlaPheThrThrThrSerProAlaLeuLeuAlaTyrLysTh 247

     696 AGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAG 745
         |||      ::::::             :::|||      :::::::::|
     247 rGlyPheProValIleAlaValAsnValSerArg......GlnAlaLysG 262

     746 GCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACAAA 795
         |||||      ::::::   |||   :::::::::|||   :::||||||
     262 lyPhe......GluValIlePro...SerAlaLysLeuTyrAlaAsnLys 275

     796 GCC............CACGATGCCGCCGTGTTCAACCGCAAT........ 825
         :::            :::::::::|||::::::      :::
     276 SerLeuProMetLysGluSerValAlaIleLeuMetAspGlnMetMetGl 292

     826 ....ACCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGT 871
             |||   |||       |||   |||:::::::::::::
     292 yPheLeuGluLysGlyIleAlaSerGlnProGluGlnTrpMetTrpIleH 309

     872 ACAACCGCTATAAA 885
         :::::|||:::|||
     309 isLysArgTrpLys 313

seq_name: sp_vertebrate:Q91238

seq_documentation_block:
ID   Q91238      PRELIMINARY;      PRT;   326 AA.
AC   Q91238;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HIGH MOLECULAR WEIGHT BASIC NUCLEAR PROTEIN (FRAGMENT).
OS   Pseudopleuronectes americanus (Winter flounder) (Pleuronectes
OS   americanus).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Pleuronectidae; Pseudopleuronectes.
OX   NCBI_TaxID=8265;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=98165790; PubMed=9497335;
RA   Watson C.E., Davies P.L.;
RT   "The high molecular weight chromatin proteins of winter flounder sperm
RT   are related to an extreme histone H1 variant.";
RL   J. Biol. Chem. 273:6157-6162(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=99196707;
RA   Watson C.E., Davies P.L.;
RT   "Recent and rapid amplification of the sperm basic nuclear protein
RT   genes in winter flounder.";
RL   Biochim. Biophys. Acta 1444:337-345(1999).
DR   EMBL; U39735; AAC13878.1; -.
KW   Nuclear protein.
FT   NON_TER       1      1
SQ   SEQUENCE   326 AA;  38103 MW;  83E3D5B773EC3F77 CRC64;


alignment_scores:

          Quality:  133.00            Length:    261
            Ratio:  1.064               Gaps:     11
Percent Similarity:  47.893   Percent Identity: 26.820

alignment_block:
 US-09-303-518D-571 x Q91238   ..

 Align seg 1/1  to: Q91238  from: 1  to: 326

     164 CGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGACACGCAG 213
         |||::::::||||||          |||||||||   ||||||      ||
      83 ArgSerLysSerPro......LysArgArgValGlnThrProLysMetAr 97

     214 ACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGC 263
         |   :::::::|||   :::|||:::|||:::   :::   ::::::   |
      97 gAlaThrSerPro...MetArgSerArgSerProSerArgTyrLysSerP 113

     264 CCCCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCGG 313
         ||   |||      |||:::   ||||||      :::   ::::::|||
     113 roLysArgSerValLysThrProLysThrArgAlaAsnSerAlaArgArg 129

     314 TACACGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTG 363

     130 ProLys..................................... 131

     364 CTGTTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACAT 413
             |||      |||:::   :::      :::   |||   :::   :
     132 ....SerProMetArgSerProSerProLysMetArgAlaLysThrArgG 147

     414 CAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGC......... 454
         ::   |||      |||||||:::   ||||||      |||
     147 lyArgSerProLysArgSerAsn...ProProLysArgSerValLysThr 162

     455 .........CGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGG.. 493
                 |||   ||||||   |||   :::||||||   :::|||
     163 ProLysThrArgAlaLysSerProArgArgSerLysSerProLysArgAr 179

     494 ............................GCAGGGTGCGCGGCAAAGGCAA 515
                                     :::      ::::::   :
     179 gValGlnThrProLysMetArgAlaLysSerProMetArgSerArgLysA 196

     516 AACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTGC 565
         ::||||||   |||:::   :::   |||          |||:::
     196 rgProArgSerProSerArgSerThrSerProMetArgSerGlnSerPro 212

     566 GCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAG 615
            |||      :::   ::::::      :::   |||||||:::|||
     213 LysArgArgValLysArgGlnLysMetThrAlaLysSerLeuMetArgSe 229

     616 GAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCAT 665
          ::::::   :::   ::::::      :::   ::::::   ||||||
     229 rArgSerProSerArgSerLysSerProLysArgSerValLysThrPro. 245

     666 GACACTGGCGGCAAAATTGGCACACGT..............CAAAGGCG 700
            |||   :::|||          |||               :::|||
     246 LysThrArgAlaLysSerProArgArgSerLysSerProLysArgArgVa 262

     701 TGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACA....... 743
          :::::::|||          ::::::||||||   |||||||:::
     262 lGlnThrProLysArgArgValGlnThrProLysArgArgValGlnThrP 279

     744 ......................AGGCTTCGTGTTGCACATCCGCCCCGT 770
                                 :::   ||||||      |||   |||
     279 roLysSerArgIleProLysProLysArgArgValProThrProLysArg 295

     771 ......CCAAGGGGAATTGAACGGCAACAAAGCCCACGATGCCGCCGTGT 814
               |||:::      :::|||   |||:::|||:::   ||||||||
     296 ArgValProLysProLysArgArgValGlnThrProLys...ArgArgVa 311

     815 TCAACCGCAATACCGAATATTGGATACGCCG 845
         ||||         |||   |||:::||||||
     311 lGlnThrProLysArgArgLeuGlnThrPro 321
```

```
seq_name: sp_bacteriap:Q9RYP2

seq_documentation_block:
ID   Q9RYP2      PRELIMINARY;      PRT;   376 AA.
AC   Q9RYP2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ADENINE DEAMINASE-RELATED PROTEIN.
GN   DRA0268.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001863; AAF12376.1; -.
DR   TIGR; DRA0268; -.
KW   Complete proteome.
SQ   SEQUENCE   376 AA;  39845 MW;  7AB7FF32F8C45651 CRC64;


alignment_scores:
           Quality:  132.00            Length:    351
             Ratio:  0.852               Gaps:     21
 Percent Similarity:  44.160   Percent Identity: 26.496

alignment_block:
 US-09-303-518D-571 x Q9RYP2   ..

 Align seg 1/1  to: Q9RYP2  from: 1  to: 376

      18 CAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCC 67
         :::|||::::::   ::::::||||||      |||         |||||
      23 ArgAlaAlaAspGluLeuGlyAsnArgCysArgHis.....ArgPr 36

      68 TGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGTCTGCACACGCT.... 113
         ||||   :::                 |||   |||::::::||||||
      36 oAla...LeuArgGlyAspAspSerAlaArgLeuAlaGlyHisAlaHisH 52

     114 ...........GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAA 152
                    |||      :::|||||||:::
      52 isThrHisPheGlyAlaAlaGlyArgThrGly.............Thr 63

     153 GGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACC 202
         |||||||||   |||                 :::|||:::   |||||
      64 GlyArgProAlaAlaGlyProAlaGlyGlyGlnSerGlyAlaGlyGluPr 80

     203 CCGA.....................CACGCAGACGGTCAAAGCCGTTTTT 231
         ||||                    ::::::   |||::::::      :
      80 oArgArgValAlaHisProAlaLeuTyrSerGlyGlyArgAlaAspArgA 97

     232 GCGGAAACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAA 281
         ::|||::::::            |||:::      |||   :::
      97 laGlyArgAlaGlySerAlaAlaGlyAlaArgTyrArgGluIleLeuGly 113

     282 ACCGGAAGACATCGAAACAAT.....................GTTCAAAG 310
           :::|||   |||::::::                     ::::::::
     114 ProAlaArgProArgGlyAspArgProTrpValSerArgArgLeuArgAl 130

     311 CGGTACACGGCTGGGAACA...CGTGCAGCAGGCTTTGGACAAGGGCGAA 357
         :::::::|||||||:::|||   |||:::||||||      :::||||||||

     130 aAlaAlaArgLeuAlaThrLeuArgGlyAlaGlyAsnLeuArgGlyArgA 147

     358 GGGCTGCTGTTCATCACGCCGCACATCGGCAG..................... 389
         ||:::::::         ||||||:::|||:::
     147 rgGlySer.........AlaAlaArgArgArgAspAlaProProAlaArg 160

     390 .....................................CTACGATTTGG 400
                                             ||||||   ::
     161 TrpArgArgArgProArgAlaArgLeuGlyArgGlyHisLeuArgSerAl 177

     401 GCGGACGCTA...CATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATG 447
         :|||:::|||   |||   |||   :::|||      ||||||||||||
     177 aArgAlaLeuAlaHisAlaProAspValSer...ArgProAspArgHisA 193

     448 TACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGC...... 491
               ::::::   :::|||      |||::::::   ||||||
     193 rgAlaCysGlyAlaGlyGln......ArgArgArgProAlaGlyArgArg 207

     492 .....................................GGGCAGGG 499
                                              ||||||
     208 ArgProLeuArgAlaArgArgValValGluLeuLeuAspValGlyGlnAr 224

     500 TGCGCGGCAAAGG............CAAAACCGCGCCCACCGGCATACAA 537
               :::|||            |||:::   |||:::|||   :::
     224 gProAlaArgArgHisProGlyHisGlnHisSerAlaArgArgAlaSerG 241

     538 GGGGTCAAACAAATCATCAA......GGCCCTGCGCGCGGG......... 572
           |||   ::::::::::|||      ||||||||||||||
     241 lyGlyProAlaArgArgGlnArgArgGlyProAlaArgGlyGlySerGly 257

     573 ............CGAGGCAACCATCATCCTGCCCGA............CC 598
                     |||   :::|||:::|||:::|||            ||
     258 AlaArgThrAlaArgArgArgHisArgProGlyArgGlyArgArgGlyPr 274

     599 ACGTCCCTTC................................TCCGCAG 615
         ||||                                    ::::::|
     274 oArgAlaValAlaProAlaLeuArgGlySerAspAspArgProAlaSerG 291

     616 GAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTGCATACACCAT 665
         |||||   ||||||      |||
     291 lyArgGlyArgArgCysProGly.....................ProGly 300

     666 GACACTGGCGGCAAAATTGGC.........ACACGTCAAAGGCGTGAAAA 706
         |||   ||||||   ::::::                    |||::::::
     301 AspGlyGlyGlyAlaLeuAlaGlyLeuHisLeuAlaLeuSerArgHisHi 317

     707 CCCTGTTTTTCTG...............................CTGC 723
         :|||   |||
     317 sProGluPheSerArgProGluArgAspSerGlyAlaGlnAlaAspAlaA 334

     724 GAACGCCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCA 773
               ||||||||||||      :::   :::|||   |||   |||
     334 laArgProAlaArgArgAspGlyValAlaIleAlaGlyProGlyArgVal 350

     774 AGG 776
         |||
     351 Arg 351

seq_name: sp_bacteria:Q53478

seq_documentation_block:
ID   Q53478      PRELIMINARY;      PRT;   456 AA.
AC   Q53478;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF2 PROTEIN.
GN   ORF2.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=95020543; PubMed=7934834;
RA   Facius D., Meyer T.F.;
RT   "A novel determinant (comA) essential for natural transformation
RT   competence in Neisseria gonorrhoeae and the effect of a comA defect on
RT   pilin variation.";
RL   Mol. Microbiol. 10:699-712(1993).
DR   EMBL; S75490; AAB32262.1; -.
SQ   SEQUENCE   456 AA;  54380 MW;  30AC7B5CDFB912D5 CRC64;


alignment_scores:
            Quality:  130.50              Length:    287
              Ratio:  0.981                 Gaps:     17
 Percent Similarity:  46.341   Percent Identity: 26.829

alignment_block:
 US-09-303-518D-571 x Q53478   ..

 Align seg 1/1  to: Q53478  from: 1  to: 456

     126 CGGACATCTGGCGTTTTACCTTTT......AAAGGAAGACCGCGC..... 164
         ||||||   |||      |||         :::   |||||||||
     170 ArgThrValGlyTyrLysProCysGlnGlnGlnHisArgProArgThrGl 186

     165 .......GCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACCCCGAC 207
                 |||   |||:::|||         :::|||   :::   :::|
     186 nThrGlnAlaAlaArgGlnGlnGlnGlnGlnSerGlyThrHisArgGlnH 203

     208 ACGCAGACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGG...... 251
         ||   :::   |||   |||      ||||||||||||:::|||
     203 isGlnGlnArgGln...ArgCysGlyGlyAsnGlyLysLeuArgArgHis 218

     252 TTTGGA.................................ACTTGCCC 265
         :::|||                                |||||
     219 IleGlyLysProHisGlnLysProGlnSerValPheArgGlnLysCysPr 235

     266 CCGCGTTTTTCAAAAAACCGGAAGACATCGAAACAATGTTCAAAGCGGTA 315
         ||||:::::::::::      ||||||   ::::::   |||
     235 oArgLeuLeuLysArgArgLysArgHisAspArgThrGlnGlnArgGlnA 252

     316 CA..............................CGGCTGGGAACACGT 332
                                        |||:::      |||
     252 rgGlyGlnHisProArgLysProArgTyrArgHisArgIleAspGlnArg 268

     333 GCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCAC........ 374
         |||          :::|||::::::   :::   ::::::
     269 AlaAspGlnArgGlnArgGlyLysGlnIleThrGluGlnAsnGlnArgGl 285

     375 ..................GCCGCACATCGGCAGCTACGATTTGGGCGGA 405
                            ||||||      |||   |||
     285 nHisArgSerArgProLeuProAlaHisGlyGlyLeuProPhe..... 299

     406 CGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCC 455
                 |||   :::::::::|||      |||      |||
     300 ..........ProLeuValGlnThrProCysThrArgProTyrGln... 311

     456 GCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCG 505
         :::::::::|||::::::||||||      :::      |||:::   ||
     312 GlyArgHisGlnThrGluArgGlnProSerThrAspThrGlnAsn...Ar 327

     506 GCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATCATC 555
         |||||||   ::::::   ||||||:::   |||
     327 gGlnArgValGluGlnGlnHisArgArgGlnArg.............L 339

     556 AAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGTCCC 605
              |||   |||   |||   :::      ||||||   |||||||||
     339 euProProArgArgProArgArgGlnSerSerProAla...ProArgPro 354

     606 TTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAACCTG 655
                   |||:::||||||
     355 ............ArgGluArgArgGln..................... 359

     656 CATACACCATGACACTGGC............GGCAAAATTGGCACACGTC 693

         |||::::::|||:::           ||||||   |||   ::::
     360 ....HisGlnAsnThrAlaLeuArgArgHisGlyLysThrGlyLysGlnA 375

     694 AAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACA 743
         ::   :::::::::               |||||||||   ||||||
     375 rgValLysGlnArg............AlaArgThrPro...ArgArgGln 386

     744 AGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCAACA 793
         :::   |||:::            :::   ||||||      :::|||
     387 AsnProArgAlaArgProGlyGlyLysAlaArgGlyThrGlyLysGlnPr 403

     794 AAGCCCACGATGCCGCCGTGTTCAACCGCAATACCGAATATTGGATACGC 843
               :::   |||:::   :::   ::::::|||   ::::::|||
     403 oLeuArgGlnProArgLysGlnHisArgHisHisArgAsnValGlnThrA 420

     844 CGTTTTCCGAC 854
               :::|||
     420 rgAsnThrAsp 423

seq_name: sp_bacteria:Q93LB1

seq_documentation_block:
ID   Q93LB1      PRELIMINARY;      PRT;   299 AA.
AC   Q93LB1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE DNA-BINDING PROTEIN.
OS   Bifidobacterium animalis (Bifidobacterium lactis).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.
OX   NCBI_TaxID=28025;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 27536;
RA   Gonzalez Vara A., Rossi M., Altomare L., Gotti R., Matteuzzi D.;
RT   "Effect of pH and culture redox potential on the continuous culture of
RT   Bifidobacterium animalis ATCC 27536, studies on stability of
RT   recombinant plasmids and characterization of the insertion sequence
RT   IS1999.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ318089; CAC41624.1; -.
KW   DNA-binding.
SQ   SEQUENCE   299 AA;  34340 MW;  B71D96C9EC0A7A03 CRC64;


alignment_scores:
            Quality:  130.00              Length:    281
              Ratio:  1.111                 Gaps:     15
 Percent Similarity:  41.637   Percent Identity: 26.335

alignment_block:
 US-09-303-518D-571 x Q93LB1   ..

 Align seg 1/1  to: Q93LB1  from: 1  to: 299

     105 GCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGG 154
         |||:::||||||:::|||:::         |||:::         :::||
      16 AlaArgAlaGlyArgProVal...LeuLeuGlyAlaArgAspArgArgGl 31

     155 AAGACCGCGC..................................... 164
         ||||   |||
      31 yArgAlaArgValAlaProArgArgValArgArgProValArgAspArgA 48

     165 .................GCGCATCGTCGCCAA................. 179
                         |||   |||||||||
      48 spArgArgProAspProAlaLeuArgArgGlnAlaGlyLysThrAlaGly 64

     180 TATGCGGCAGGCGGGTTTGAACCCCGACACGCAGACGGTCAAAGCCGTTT 229
         ::::::||||||   |||:::|||
      65 HisGlyAlaGlyProPheArgPro..................... 72

     230 TTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCCCCGCGTTTTT.... 275
```

```
          |||   ::::::|||      |||:::      ||||||
 73 ....GlyAlaSerArgMetAspAlaGlyAlaArgArgArgValHisLysT 88

276 ..CAAAAAACCGGAAGACATCGA......AACAATGTTCAAAGCGGTACA 317
      ::::::   |||||||||         ::::::::::::::|||'
 88 hrArgArgGlyGlyArgHisGlyArgValArgArgLeuGluAlaGlyArg 104

318 CGGCTGGGAACACGTGCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGT 367
    |||:::      |||   |||      |||:::|||:::|||   :::
105 ArgIle......ArgArgAlaProArgGlyArgGlyHisArgProValPr 119

368 TCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGC 417
     ::::::|||:::   |||:::   |||                ||||||
119 oArgArgAlaThrGlyArgArgProArg..............HisGlnG 131

418 CAGCAGCTTCCGTTCCACCTGACCGCCATGTACAAGCCGCCGAAAATCAA 467
      |||                 ||||||  ||||||
131 lyAlaLeuProProAlaThrArgArgHisArgGlnAla............ 143

468 AGCGATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGG..... 512
                :::      |||   |||:::   |||   :::
144 ............ArgCysGlnGlyArgGlnAlaLeuArgLeuGlnAlaAr 156

513 ..........CAAAACCGCGCCCACCGGCATACAAGGGGTCAAACAAATC 552
              |||:::|||      |||:::||||||         :::|
156 gProAlaAspGlnArgArgIleProArgArgThrArgProCysGlnThrH 173

553 ATCAAGGCCCTGCGCGCGGGCGAGGCAACCATCATCCTGCCCGACCACGT 602
    ||   ||||||:::|||                ||||||::::::||||||
173 is...GlyProValArg..............HisProGlyGlnProArg 183

603 CCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTTTCGGCAAAC 652
    |||   :::      ||||||         |||
184 ProHisAlaArgProArgArgLeuProLysGly................. 194

653 CTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTCAAAGGCGTG 702
              |||
195 .........His..................................... 195

703 AAAACCCTGTTTTTCTGCTGCGAACGCCTGCCCGACGGACAAGGCTTCGT 752
                      |||   |||:::||||||         |||
196 ................ThrArgLeuProValArgArgGlnAlaThrArg 206

753 GTTGCACAT.....................CCGCCCCGTCCAAGGGGAAT 781
      :::|||                     |||   |||      :::::
207 ArgThrHisAspAlaArgAlaAspArgArgProAspArgThrGlyAlaLe 223

782 TGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAA 824
    ::::|||      :::   |||      |||   :::   |||
223 uGlnArgLeuProGlyThrArgLeuProArgProArgAlaGln 237
```

```
seq_name: sp_bacteria:O85018

seq_documentation_block:
ID   O85018      PRELIMINARY;      PRT;  1122 AA.
AC   O85018;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXIT.
GN   EXIT.
OS   Mycobacterium smegmatis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1772;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LR222;
RX   MEDLINE=98385832; PubMed=9720878;
RA   Zhu W.M., Arceneaux J.E.L., Beggs M.L., Byers B.R., Eisenach K.D.,
RA   Lundrigan M.D.;
RT   "Exochelin genes in Mycobacterium smegmatis: identification of an ABC
RT   transporter and two non-ribosomal peptide synthetase genes.";
RL   Mol. Microbiol. 29:629-639(1998).
RN   [2]
RP   SEQUENCE OF 534-1122 FROM N.A.
RC   STRAIN=MC2155;
RX   MEDLINE=95191405; PubMed=7885234;
RA   Fiss E.H., Yu S., Jacobs W.R. Jr.;
RT   "Identification of genes involved in the sequestration of iron in
RT   mycobacteria: the ferric exochelin biosynthetic and uptake pathways.";
RL   Mol. Microbiol. 14:557-569(1994).
RN   [3]
RP   SEQUENCE OF 534-1122 FROM N.A.
RC   STRAIN=MC2155;
RX   MEDLINE=98389687; PubMed=9721311;
RA   Yu S., Fiss E., Jacobs W.R. Jr.;
RT   "Analysis of the exochelin locus in Mycobacterium smegmatis:
RT   biosynthesis genes have homology with genes of the peptide synthetase
RT   family.";
RL   J. Bacteriol. 180:4676-4685(1998).
DR   EMBL; AF034152; AAC32046.1; -.
DR   EMBL; AF027770; AAC82548.1; -.
DR   InterPro; IPR001140; ABC_transporter_tmem.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00664; ABC_membrane; 1.
DR   Pfam; PF00005; ABC_tran; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1122 AA;  120386 MW;  4EC826C67914A57A CRC64;


alignment_scores:
            Quality:  130.00              Length:    337
              Ratio:  0.828                Gaps:     18
 Percent Similarity:  46.588   Percent Identity: 26.409

alignment_block:
 US-09-303-518D-571 x O85018   ..

 Align seg 1/1  to: O85018  from: 1  to: 1122

       3 GTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACA 52
         |||   |||   :::|||   ::::::|||:::      |||::::::::
     135 ValValPheGlyLeuGlnLeuThrAlaProMetIleAlaArgArgValGl 151

      53 TCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGTCGCTTTCCTGT 102
         :      |||:::   |||::::::
     151 u......AspGlnGlnAlaGluIleGly.....................A 159

     103 CTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAA 152
           |||   |||                |||||      |||::::::
     159 rgAlaThrAlaLeuAlaThrAspLeuIleSerGlyGlnArgProLeuGln 175

     153 GGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGGCGGGTTTGAACC 202
         |||:::   ||||||   :::|||      |||            |||||
     176 GlyAsnArgArgAlaGlyGlnArgGlyArgAlaLeuProCysArgGluPr 192

     203 CCGACACGCAGAC..................................... 215
         |   :::||||||
     192 oSerArgAlaAspArgAspValAlaCysArgAlaHisProGluHisProL 209

     216 ..GGTCAAAGCCGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACT... 260
           |||:::         ||||||:::|||:::         ||||||
     209 euGlyArg.........CysGlyGlyGlyArg.........GlyThrArg 219

     261 .................................TGCCCCCGCGTTTT 274
                                          ||||||   ||
     220 ArgAspGlyGlyGlyArgGlyArgGlyLeuLeuArgAspProArg...Ph 235

     275 TCAAAAAACCGGA..............................AGAC 291
         |:::      |||                             |
     235 eAspAspGlyGlyProValAspHisArgAspArgProArgProValProH 252

     292 ATCGAAACAATGTTCAAAGCGGTACACGG......CTGGGAACACGTGCA 335
         ||||||:::   :::   :::|||   |||      ||||||:::|||:::
```

```
    252 isArgThrValLeuAlaAlaGlyGlyArgAlaGluLeuGlySerArgGly 268

    336 GCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCG 385
        |||   ||||||   |||||||:::      ::::::
    269 ProGlyValGlyGlnProArgArgSerArgProGluArgProAspProPr 285

    386 GCAGCTACGATTT...GGGCGGACGCTACATCAGCCAGCAGCTTCCGTTC 432
            |||:::   |||         ::::::   :::      |||
    285 oLeuProArgLeuArgGlyGlyProProArgArgArgGlyProProValG 302

    433 CACCTGACCGCCATGTACAAGCCGCCGAA.....AATCAAAGC 470
        |||:::|||:::::::::::::::::|||         ::::::
    302 lyProGlnArgAsnThrArgGlySerGluLysProValValArgArgArg 318

    471 GATAGACAAAATCATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCG 520
        |||:::   :::   |||   :::      |||   |||   :::||
    319 AlaArgArgIleArgArgGlyThrHisGluArgArgProArgCysArgAr 335

    521 CGCCCACCGGCAT.................ACAAGGGGTCAAACAAATC 552
        |:::||||||                    :::|||||||||   :::
    335 gThrHisArgSerAlaLeuArgIleSerAlaSerArgGlyGlnArgHisG 352

    553 ATCAAGGC.....................CCTGCGCGCGGGCGAGGC 578
            |||                      ||||||:::   |||:::
    352 lyAlaGlyArgTrpProAlaProArgProAspProAlaGlu...ArgAla 367

    579 AACCATCATCCTGCCCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCG 628
        :::   |||||||||      |||   :::      :::|||
    368 ProArgThrProAlaArgArgThrProProGlnLeuValValGlnArgAs 384

    629 TGTGGGCGGATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCA 678
        :::|||:::      ::::::   |||   ::::::   ::::::::::
    384 pAlaGlyAsn......GlnSerGluCysHisArgAsnArgSerAlaAlaA 399

    679 AAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACG 728
        ::   |||   |||:::||||||   :::   |||
    399 rgThrGlyLeuArgArgArgArgThrThrGlyVal.............. 410

    729 CCTGCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGG 778
        :::||||||   |||   |||   |||   |||   |||   |||
    411 ...ValArgArgArgArgArgArgProAla...ProArgArgSerArgPr 425

    779 AATTGAACGGCAACAAAGCCCACGATGCCGCCGTGTTCAACCGCAATACC 828
        :::|||      :::   ||||||:::   |||:::
    425 oCysArgArg......GlyAlaArgCysGluProValArgArgThrThrP 440

    829 GAATATTGGAT 839
            ||||||
    440 roThrLeuAsp 443

seq_name: sp_virus:Q9QCQ3

seq_documentation_block:
ID   Q9QCQ3      PRELIMINARY;      PRT;   297 AA.
AC   Q9QCQ3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   THYMIDINE KINASE.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10345;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PRV-SH;
RA   Zhu C.Q., Wang H.J., Qi X., Shen Q., Ye X.Y., Chen P.Y., Cai B.X.;
RT   "Cloning and sequence analysis of the thymidine kinase gene from
RT   pseudorabies virus (PRV-SH) isolated in Shanghai, China.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PRV-SH;
RA   Pan J.Y., Chen D.S., Chen P.Y., Cai B.X.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF199413; AAF06353.1; -.
DR   HSSP; P03176; 2VTK.
DR   InterPro; IPR001889; TK_herpes.
DR   Pfam; PF00693; TK_herpes; 1.
DR   ProDom; PD001519; TK_herpes; 1.
KW   Kinase.
SQ   SEQUENCE   297 AA;  31630 MW;  D8C36A4694AEB465 CRC64;


alignment_scores:
            Quality:  128.50              Length:    164
              Ratio:  1.530                 Gaps:      9
 Percent Similarity:  51.220   Percent Identity: 29.878

alignment_block:
 US-09-303-518D-571 x Q9QCQ3   ..

 Align seg 1/1  to: Q9QCQ3  from: 1  to: 297

     245 AATGCGGTTTGGAACTTGCCCCCGCGTTTTTCAAAAAACCGGAAGACATC 294
         |||   ||||||          ||||||:::::::::::::   :::|||
     148 AsnLeuValTrp.........ProArgTrpThrArgThr...SerThrCy 160

     295 GAAACAATGTTCAAAGCGGTACACGGCTGGGAACACG.............T 332
               |||:::         :::   |||:::|||             |
     160 sGlyAlaCysAlaProAlaArgAlaProGlySerThrTrpThrArgAlaC 177

     333 GCAGCAGGCTTTGGACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACA 382
         ||||||||      :::   :::      |||   |||:::||||||:::
     177 ysSerArgAlaCysAlaThrSerThrProCysTrpSerThrArgArgAla 193

     383 TCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTC 432
         :::   |||         |||   ||||||:::::::::::::   |||::
     194 ThrLeuAlaArgGlyAlaAlaGlyAlaThrThrGlyGlyAlaArgArgAl 210

     433 CACCTGACCGCCATGTACAAGCCGCCGAAAATCAAAGCGATAGACAAAAT 482
         ::::      ||||||::::::      :::|||         |||:::|
     210 aSerThrArgProCysAlaThrAlaSerArgSer.........ThrSerS 224

     483 CATGCAGGCGGGCAGGGTGCGCGGCAAAGGCAAAACCG............ 520
         ||:::   ||||||   :::   ::::::      |||
     224 erAlaAlaArgAlaThrAlaProSerSerArgThrProSerSerAlaArg 240

     521 ......CGCCCACCGGCATACAAGGGGTCAAACAAATCATCAAGGCCCTG 564
               ||||||   :::      |||::::::      ||||||   ||
     241 ThrArgArgProSerSerAlaThrGlyAlaGlyAlaArgSerArg...Cy 256

     565 CGCGCGGGCGAGG............CAACCATCATCCTGC.......... 592
         |:::|||:::|||              ||||||   |||
     256 sThrArgGlyArgTrpThrArgSerTrpProSerCysCysArgCysAlaS 273

     593 .....CCGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGG 637
              |||:::|||      |||||||||   :::|||
     273 erProProSerThrTrpGlyLeuArgArgAlaSerAla............ 285

     638 ATTTTTTCGGCAAACCTGCATACACCATGACACTGGCGGCAA 679
                                 |||      ||||||:::
     286 .......................ProArgProTrpArgArg 291

seq_name: sp_human:Q9UQ39

seq_documentation_block:
ID   Q9UQ39      PRELIMINARY;      PRT;   956 AA.
AC   Q9UQ39;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA BINDING PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016088; BAA83714.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER       1      1
FT   NON_TER     956    956
SQ   SEQUENCE   956 AA;  106462 MW;  C9FEAD07B49F5011 CRC64;


alignment_scores:
            Quality:  128.00             Length:    260
              Ratio:  1.153                Gaps:     11
 Percent Similarity:  42.692   Percent Identity: 26.154

alignment_block:
 US-09-303-518D-571 x Q9UQ39   ..

 Align seg 1/1  to: Q9UQ39  from: 1  to: 956

     152 AGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGG.......... 190
         |||:::   :::|||:::   ||||||      ::::::
     444 ArgArgMetGlyArgSerArgSerProAlaThrAlaLysArgGlyArgSe 460

     191 .CGGGTTTGAACCCCGACACGCAGA..............CGGTCAAAGC 224
          |||      |||||||||:::|||               ||||||
     460 rArgSerArgThrProThrLysArgGlyHisSerArgSerArgSerProG 477

     225 CGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCC......... 265
           :::   |||:::|||:::         |||:::
     477 lnTrpArgArgSerArgSerArgSerAlaGlnArgTrpGlyArgSerArgSerPro 493

     266 ..............................CCGCGTTTTTCAAAAAA 282
                                       |||   :::|||:::::
     494 GlnArgArgGlyArgSerArgSerProGlnArgProGlyTrpSerArgSe 510

     283 CCGGAAGACATCGAAACAATGTTCAAAGCGG................... 313
         :|||:::|||:::::::::      :::|||
     510 rArgAsnThrGlnArgArgGlyArgSerArgSerAlaArgArgGlyArgS 527

     314 .................TACACGGCTGGGAACACGTGCAGCAGGCTTTGG 346
                           |||   |||::::::   ||||||
     527 erHisSerArgSerProAlaThrArgGlyArgSerArgSerArg...... 541

     347 ACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGAT 396
         |||   |||:::        |||      ||||||   |||
     542 ThrProAlaArgArgGlyArgSerArgSerArgThrProAla........ 555

     397 TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCAT 446

     555 .................................................. 555

     447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCA 496
                |||||:::|||:::      |||   :::   ||||||
     556 ........ArgArgArgSerArgSerArgThrProThrArgArgArg... 568

     497 GGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAA 546
                       ::::::   |||   ||||||   :::|||   ::
     569 ..............SerArgSerArgThrProAlaArgArgGlyArgSe 580

     547 CAAATCATCAAGGCCCTGCGCGCGGGCGAGGC...AACCATCATCCTGCC 593
         ::::||| :::   |||||||||   |||:::   ::::::   |||:::
     580 rArgSer.ArgThrProAlaArgArgArgSerArgThrArgSerProVal 596

     594 CGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTT 643
         |||   |||      |||   :::      ||||||   |||
     597 ArgArgArgSerArgSerArgSerProAlaArgArgSerGly........ 610

     644 TCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTC 693
```

```
                                                        |||:
     611 .............................................ArgS 612

     694 AAAGGCGTGAAAACCCTGTT............TTTCTGCTGCGAACGCCT 731
         ::|||   ::::::|||:::                |||||||||
     612 erArgSerArgThrProAlaArgArgGlyArgSerArgSerArgThrPro 628

     732 GCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAAT 781
         |||||||||   |||   |||          |||   |||      |||
     629 AlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSerGlyAr 645

     782 TGAACGGCAACAAAGCCCACGATGCCGC 809
          :::|||:::::::::|||      |||
     645 gSerArgSerArgThrProAlaArgArg 654

seq_name: sp_human:Q9UQ40

seq_documentation_block:
ID   Q9UQ40      PRELIMINARY;      PRT;  1262 AA.
AC   Q9UQ40;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA BINDING PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016087; BAA83713.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER       1      1
FT   NON_TER    1262   1262
SQ   SEQUENCE   1262 AA;  138003 MW;  75E76BD73FA14295 CRC64;


alignment_scores:
            Quality:  128.00             Length:    260
              Ratio:  1.153                Gaps:     11
 Percent Similarity:  42.692   Percent Identity: 26.154

alignment_block:
 US-09-303-518D-571 x Q9UQ40   ..

 Align seg 1/1  to: Q9UQ40  from: 1  to: 1262

     152 AGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGG.......... 190
         |||:::   :::|||:::   ||||||      ::::::
     269 ArgArgMetGlyArgSerArgSerProAlaThrAlaLysArgGlyArgSe 285

     191 .CGGGTTTGAACCCCGACACGCAGA..............CGGTCAAAGC 224
          |||      |||||||||:::|||               ||||||
     285 rArgSerArgThrProThrLysArgGlyHisSerArgSerArgSerProG 302

     225 CGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCC......... 265
           :::   |||:::|||:::         |||:::
     302 lnTrpArgArgSerArgSerArgSerAlaGlnArgTrpGlyArgSerArgSerPro 318

     266 ..............................CCGCGTTTTTCAAAAAA 282
                                       |||   :::|||:::::
     319 GlnArgArgGlyArgSerArgSerProGlnArgProGlyTrpSerArgSe 335

     283 CCGGAAGACATCGAAACAATGTTCAAAGCGG................... 313
         :|||:::|||:::::::::      :::|||
     335 rArgAsnThrGlnArgArgGlyArgSerArgSerAlaArgArgGlyArgS 352

     314 .................TACACGGCTGGGAACACGTGCAGCAGGCTTTGG 346
                           |||   |||::::::   ||||||
```

```
     352 erHisSerArgSerProAlaThrArgGlyArgSerArgSerArg..... 366

     347 ACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGAT 396
         |||   |||:::        |||      ||||||   |||
     367 ThrProAlaArgArgGlyArgSerArgSerArgThrProAla..... 380

     397 TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCAT 446

     380 ..................................................... 380

     447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCA 496
                 ||||||:::|||:::      |||   :::   ||||||
     381 ........ArgArgArgSerArgSerArgThrProThrArgArgArg... 393

     497 GGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAA 546
                       ::::::   |||   ||||||   :::|||   ::
     394 ..............SerArgSerArgThrProAlaArgArgGlySe 405

     547 CAAATCATCAAGGCCCTGCGCGCGGGCGAGGC...AACCATCATCCTGCC 593
         :::|||  :::   |||||||||   |||:::   ::::::   |||:::
     405 rArgSer.ArgThrProAlaArgArgArgSerArgThrArgSerProVal 421

     594 CGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTT 643
         |||   |||      |||   :::      ||||||   |||
     422 ArgArgArgSerArgSerArgSerProAlaArgArgSerGly........ 435

     644 TCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTC 693
                                                    |||:
     436 ..............................................ArgS 437

     694 AAAGGCGTGAAAACCCTGTT............TTTCTGCTGCGAACGCCT 731
         ::|||   ::::::|||:::                  |||||||||
     437 erArgSerArgThrProAlaArgArgGlyArgSerArgSerArgThrPro 453

     732 GCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAAT 781
         |||||||||   |||   |||         |||   |||      |||
     454 AlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSerGlyAr 470

     782 TGAACGGCAACAAAGCCCACGATGCCGC 809
          :::|||:::::::::|||      |||
     470 gSerArgSerArgThrProAlaArgArg 479

seq_name: sp_human:Q9UHA8

seq_documentation_block:
ID   Q9UHA8      PRELIMINARY;      PRT;  2296 AA.
AC   Q9UHA8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SPLICING COACTIVATOR SUBUNIT SRM300.
GN   SRM300.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20132238; PubMed=10668804;
RA   Blencowe B.J., Bauren G., Eldridge A.G., Issner R., Nickerson J.A.,
RA   Rosonina E., Sharp P.A.;
RT   "The SRm160/300 splicing coactivator subunits.";
RL   RNA 6:111-120(2000).
DR   EMBL; AF201422; AAF21439.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   2296 AA;  251964 MW;  17C0BD4EA10A9CF9 CRC64;


alignment_scores:
           Quality:  128.00            Length:    260
             Ratio:  1.153               Gaps:     11
 Percent Similarity:  42.692   Percent Identity: 26.154
```

```
alignment_block:
 US-09-303-518D-571 x Q9UHA8   ..

 Align seg 1/1  to: Q9UHA8  from: 1  to: 2296

     152 AGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGG..... 190
         |||:::   :::|||:::   ||||||      ::::::
     479 ArgArgMetGlyArgSerArgSerProAlaThrAlaLysArgGlyArgSe 495

     191 .CGGGTTTGAACCCCGACACGCAGA.............CGGTCAAAGC 224
          |||      |||||||||:::|||               ||||||
     495 rArgSerArgThrProThrLysArgGlyHisSerArgSerArgSerProG 512

     225 CGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCC..... 265
           :::   |||:::|||:::         |||:::
     512 lnTrpArgArgSerArgSerAlaGlnArgTrpGlyArgSerArgSerPro 528

     266 ..............................CCGCGTTTTTCAAAAAA 282
                                          |||   :::|||:::::
     529 GlnArgArgGlyArgSerArgSerProGlnArgProGlyTrpSerArgSe 545

     283 CCGGAAGACATCGAAACAATGTTCAAAGCGG................... 313
         :|||:::|||:::::::::      :::|||
     545 rArgAsnThrGlnArgArgGlyArgSerArgSerAlaArgArgGlyArgS 562

     314 ................TACACGGCTGGGAACACGTGCAGCAGGCTTTGG 346
                              |||   |||::::::   ||||||
     562 erHisSerArgSerProAlaThrArgGlyArgSerArgSerArg..... 576

     347 ACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGAT 396
         |||   |||:::        |||      ||||||   |||
     577 ThrProAlaArgArgGlyArgSerArgSerArgThrProAla..... 590

     397 TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCAT 446

     590 ..................................................... 590

     447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCA 496
                 ||||||:::|||:::      |||   :::   ||||||
     591 ........ArgArgArgSerArgSerArgThrProThrArgArgArg... 603

     497 GGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAA 546
                       ::::::   |||   ||||||   :::|||   ::
     604 ..............SerArgSerArgThrProAlaArgArgGlyArgSe 615

     547 CAAATCATCAAGGCCCTGCGCGCGGGCGAGGC...AACCATCATCCTGCC 593
         :::|||  :::   |||||||||   |||:::   ::::::   |||:::
     615 rArgSer.ArgThrProAlaArgArgArgSerArgThrArgSerProVal 631

     594 CGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTT 643
         |||   |||      |||   :::      ||||||   |||
     632 ArgArgArgSerArgSerArgSerProAlaArgArgSerGly........ 645

     644 TCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTC 693
                                                    |||:
     646 ..............................................ArgS 647

     694 AAAGGCGTGAAAACCCTGTT............TTTCTGCTGCGAACGCCT 731
         ::|||   ::::::|||:::                  |||||||||
     647 erArgSerArgThrProAlaArgArgGlyArgSerArgSerArgThrPro 663

     732 GCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAAT 781
         |||||||||   |||   |||         |||   |||      |||
     664 AlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSerGlyAr 680

     782 TGAACGGCAACAAAGCCCACGATGCCGC 809
          :::|||:::::::::|||      |||
     680 gSerArgSerArgThrProAlaArgArg 689

seq_name: sp_human:Q9UQ35

seq_documentation_block:
```

```
ID   Q9UQ35      PRELIMINARY;      PRT;  2752 AA.
AC   Q9UQ35;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA BINDING PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, complete cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016092; BAA83718.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   2752 AA;  299672 MW;  109C64F181097123 CRC64;


alignment_scores:
            Quality:  128.00              Length:    260
              Ratio:  1.153                 Gaps:     11
 Percent Similarity:  42.692   Percent Identity: 26.154

alignment_block:
 US-09-303-518D-571 x Q9UQ35   ..

 Align seg 1/1  to: Q9UQ35  from: 1  to: 2752

     152 AGGAAGACCGCGCGCGCATCGTCGCCAATATGCGGCAGG..........  190
         |||:::   :::|||:::   ||||||      ::::::
     479 ArgArgMetGlyArgSerArgSerProAlaThrAlaLysArgGlyArgSe  495

     191 .CGGGTTTGAACCCCGACACGCAGA.............CGGTCAAAGC  224
          |||      ||||||||||:::|||               ||||||
     495 rArgSerArgThrProThrLysArgGlyHisSerArgSerArgSerProG  512

     225 CGTTTTTGCGGAAACGGCAAAATGCGGTTTGGAACTTGCCC.........  265
           :::   |||:::|||:::         |||:::
     512 lnTrpArgArgSerArgSerAlaGlnArgTrpGlyArgSerArgSerPro  528

     266 ...............................CCGCGTTTTTCAAAAAA  282
                                        |||   :::|||:::::
     529 GlnArgArgGlyArgSerArgSerProGlnArgProGlyTrpSerArgSe  545

     283 CCGGAAGACATCGAAACAATGTTCAAAGCGG...................  313
         :|||:::|||:::::::::      :::|||
     545 rArgAsnThrGlnArgArgGlyArgSerArgSerAlaArgArgGlyArgS  562

     314 ................TACACGGCTGGGAACACGTGCAGCAGGCTTTGG  346
                         |||   |||::::::   ||||||
     562 erHisSerArgSerProAlaThrArgGlyArgSerArgSerArg.....  576

     347 ACAAGGGCGAAGGGCTGCTGTTCATCACGCCGCACATCGGCAGCTACGAT  396
         |||   |||:::         |||      ||||||   |||
     577 ThrProAlaArgArgGlyArgSerArgSerArgThrProAla........  590

     397 TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCACCTGACCGCCAT  446

     590 ..................................................  590

     447 GTACAAGCCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCA  496
                |||||||:::|||:::      |||   :::   ||||||
     591 ........ArgArgArgSerArgSerArgThrProThrArgArgArg...  603

     497 GGGTGCGCGGCAAAGGCAAAACCGCGCCCACCGGCATACAAGGGGTCAAA  546
                       ::::::   |||   ||||||   :::|||   ::
     604 ..............SerArgSerArgThrProAlaArgArgGlyArgSe  615

     547 CAAATCATCAAGGCCCTGCGCGCGGGCGAGGC...AACCATCATCCTGCC  593
         ::::||| :::   |||||||||   |||:::   ::::::   |||:::

     615 rArgSer.ArgThrProAlaArgArgArgSerArgThrArgSerProVal  631

     594 CGACCACGTCCCTTCTCCGCAGGAAGGCGGCGGCGTGTGGGCGGATTTTT  643
         |||   |||       |||   :::       ||||||   |||
     632 ArgArgArgSerArgSerArgSerProAlaArgArgSerGly........  645

     644 TCGGCAAACCTGCATACACCATGACACTGGCGGCAAAATTGGCACACGTC  693
                                                       |||:
     646.....................................................ArgS  647

     694 AAAGGCGTGAAAACCCTGTT.............TTTCTGCTGCGAACGCCT  731
         ::|||   :::::::|||:::                      |||||||||
     647 erArgSerArgThrProAlaArgArgGlyArgSerArgSerArgThrPro  663

     732 GCCCGACGGACAAGGCTTCGTGTTGCACATCCGCCCCGTCCAAGGGGAAT  781
         |||||||||   |||   |||         |||   |||       |||
     664 AlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSerGlyAr  680

     782 TGAACGGCAACAAAGCCCACGATGCCGC  809
          :::|||::::::::::|||      |||
     680 gSerArgSerArgThrProAlaArgArg  689
```